

New Methods for Inference in Long-Run Predictive Regressions

K.7

Erik Hjalmarsson



International Finance Discussion Papers

Board of Governors of the Federal Reserve System

Number 853

January 2006

Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 853

January 2006

New Methods for Inference in Long-Run Predictive Regressions

Erik Hjalmarsson

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDPs are available on the Web at www.federalreserve.gov/pubs/ifdp/.

New Methods for Inference in Long-run Predictive Regressions*

Erik Hjalmarsson[†]

Division of International Finance

Federal Reserve Board, Mail Stop 20, Washington, DC 20551, USA

January 2006

Abstract

I develop new asymptotic results for long-horizon regressions with overlapping observations. I show that rather than using auto-correlation robust standard errors, the standard t-statistic can simply be divided by the square root of the forecasting horizon to correct for the effects of the overlap in the data. Further, when the regressors are persistent and endogenous, the long-run OLS estimator suffers from the same problems as does the short-run OLS estimator, and similar corrections and test procedures as those proposed for the short-run case should also be used in the long-run. In addition, I show that under an alternative of predictability, long-horizon estimators have a slower rate of convergence than short-run estimators and their limiting distributions are non-standard and fundamentally different from those under the null hypothesis. These asymptotic results are supported by simulation evidence and suggest that under standard econometric specifications, short-run inference is generally preferable to long-run inference. The theoretical results are illustrated with an application to long-run stock-return predictability.

JEL classification: C22, G1.

Keywords: Predictive regressions; Long-horizon regressions; Stock return predictability.

*This paper was part of the author's Ph.D. dissertation at Yale University and parts of the contents were previously circulated under the title "On the Predictability of Global Stock Returns".

[†]I have greatly benefitted from comments by my advisors, Peter Phillips and Robert Shiller. I am also grateful for advice from Don Andrews, Ray Fair, Jon Faust, Lennart Hjalmarsson, Randi Hjalmarsson, Yuichi Kitamura, Taisuke Otsu, as well participants in the econometrics seminar and workshop at Yale University, the finance seminar at Göteborg University, and the World meeting of the Econometric Society in London. Tel.: +1-202-452-2436; fax: +1-202-263-4850; email: erik.hjalmarsson@frb.gov. The views presented in this paper are solely those of the author and do not represent those of the Federal Reserve Board or its staff.

1 Introduction

Predictive regressions are used frequently in empirical finance and economics. The underlying economic motivation is often the test of a rational expectations model, which implies that the innovations to the dependent variable should be orthogonal to all past information; i.e., the dependent variable should not be predictable using any lagged regressors. Although this orthogonality condition should hold at any time horizon, it is popular to test for predictability by regressing sums of future values of the dependent variable onto the current value of the regressor. A leading example is the question of stock return predictability, where regressions with 5 or 10 year returns are often used (e.g. Campbell and Shiller, 1988, and Fama and French, 1988a); while stock return predictability will also serve as the motivating example in this paper, the results derived are applicable to a much wider range of empirical questions.¹

The reasons behind the use of such long-horizon regressions are seldom clearly spelled out, but there seems to be an intuitive notion that a stronger signal might be extracted by analyzing long-run data. Previous analytical evidence in favour of this intuition is somewhat mixed, and overall fairly weak (e.g. Berkowitz and Giorgianni, 2001, Campbell, 2001, Mark and Sul, 2004, and Rapach and Wohar, 2005). From a practical perspective, however, the main issue with long-horizon regressions has been the uncertainty regarding the proper calculation of standard errors. Since overlapping observations are typically used, the regression residuals will exhibit strong serial correlation and standard errors failing to account for this fact will lead to biased inference.

The main contribution of this paper is the development of new asymptotic results for long-run regressions with overlapping observations. Using a framework with near-unit-root regressors, I show how to obtain asymptotically correct test-statistics, with good small sample properties, for the null hypothesis of no predictability. In addition, I provide some further understanding of long-horizon regressions, and the power of the associated test-statistics, by analyzing the properties of the long-run estimators under an alternative of predictability.²

¹Other applications of long-horizon regressions include tests of exchange rate predictability (Mark, 1995, and Rossi 2005), the Fisher effect (Mishkin, 1990, 1992, and Boudoukh and Richardson, 1993), and the neutrality of money (Fisher and Seater, 1993).

²There is now a large literature on regressions with overlapping observations. Additional references to those mentioned previously include Hansen and Hodrick (1980), Richardson and Stock (1989), Richardson and Smith (1991), Hodrick (1992), Nelson and Kim (1993), Goetzman and Jorion (1993), Daniel (2001), Moon et al. (2004), Torous et al. (2004), and Boudoukh et al. (2005). The study by Valkanov (2003) is the most closely related to this paper and is discussed in more detail below.

Studies on (short-run) predictive regressions in the context of persistent regressors include Mankiw and Shapiro (1986), Cavanagh et al. (1995), Stambaugh (1999), Lewellen (2004), Janson and Moreira (2004), Polk et al. (2004), and

Typically, auto-correlation robust estimation of the standard errors (e.g. Newey and West, 1987) is used to perform inference in long-run regressions. However, these robust estimators tend to perform poorly in finite samples since the serial correlation induced in the error terms by overlapping data is often very strong. I show that rather than using robust standard errors, the standard t -statistic can simply be divided by the square root of the forecasting horizon to correct for the effects of the overlap in the data. Further, when the regressors are persistent and endogenous, the long-run OLS estimator suffers from the same problems as does the short-run OLS estimator, and similar corrections and test procedures as those proposed by Campbell and Yogo (2005) for the short-run case should also be used in the long-run; again, the resulting test statistics should be scaled due to the overlap. Thus, these results lead to simple and more efficient inference in long-run regressions by obviating the need for robust standard error estimation methods and controlling for the endogeneity and persistence of the regressors.

The asymptotic distributions of the long-run estimators are derived not only under the null-hypothesis of no predictability, but also under an alternative of predictability. This gives a more complete characterization of the asymptotic properties of the long-run estimators than is typically found in the literature, where results for long-run estimators are often derived only under the null-hypothesis of no predictability. It is shown that, under the standard econometric model of stock return predictability, the long-run estimators converge to well-defined quantities, but their asymptotic distributions are non-standard and fundamentally different from the asymptotic distributions under the null hypothesis of no predictability. The rates of convergence of the long-run estimators are also slower under the alternative hypothesis of predictability than under the null hypothesis, and slower than that of the short-run estimator. These results suggest that under the standard econometric specifications that are typically postulated, short-run inference is preferable to long-run inference. These results are also in line with those of Mark and Sul (2004), who analyze local alternatives to the null hypothesis of no predictability. They find that there are cases in which a long-run specification has more power to detect deviations from the null hypothesis than a short-run specification, but only when the innovations to the regressand are correlated with *past* innovations to the regressor. Since such a correlation already violates the orthogonality condition that is being tested, this region of the parameter space is not relevant in most traditional applications of long-horizon regressions.

The results in this paper are derived under the assumption that the forecasting horizon increases

Campbell and Yogo (2005).

with the sample size, but at a slower pace. Most previous work, e.g. Richardson and Stock (1989) and Valkanov (2003), rely on the assumption that the forecasting horizon grows at the same pace as the sample size so that the forecasting horizon remains a fraction of the sample size asymptotically. In some related work, Moon et al. (2004) consider both asymptotic approaches and find that although the asymptotic distributions are different under the two assumptions, they both tend to provide good approximations for the finite sample properties. Indeed, Valkanov (2003), who studies a similar econometric model to the one analyzed in this paper, derives a similar scaling result to the one found here. His results are less transparent, however, and do not capture the convenient result derived here — namely, that with exogenous regressors the scaled standard t -statistic will be normally distributed. Likewise, Valkanov's results are also not as amenable to the endogeneity corrections pursued in this paper.

Monte-Carlo simulations confirm the asymptotic results just described. They show that the scaling of the test-statistic to control for the overlap in the observations, in conjunction with the endogeneity corrections, produce tests that have rejection rates under the null hypothesis very close to their nominal size. Further, it is evident that the power of the tests, in finite samples, based on short-run inference dominates that of the long-run test-statistics, which confirms the asymptotic results under the alternative of predictability.

The theoretical results in the paper are illustrated with an application to stock-return predictability. I use annual excess returns on the S&P 500, as well as monthly excess returns on the value weighted CRSP index as dependent variables, and the dividend price ratio and the smoothed earnings price ratio, suggested by Campbell and Shiller (1988), as predictor variables. The data are a subset of those used by Campbell and Yogo (2005) and for the one period horizon case the results are qualitatively identical to those found by Campbell and Yogo. Using the entire samples and correcting for the endogeneity and persistence in the regressors, the earnings price ratio is found to be a significant predictor in both the annual and the monthly data, whereas the null of no predictability cannot be rejected when using the dividend price ratio as a predictor. Contrary to some popular beliefs, however, the case for predictability does not increase with the forecast horizon. In fact, the near opposite is true and for horizons beyond five years the null of no predictability can no longer be rejected for the earnings price ratio. Given the fairly weak evidence of predictability at the short horizon, these results are entirely consistent with a loss of power as the forecasting horizon increases.

The rest of the paper is organized as follows. Section 2 sets up the model and the key assumptions. Section 3 derives the theoretical results and Section 4 describes the Monte-Carlo simulations. Section 5 discusses the practical implementation of the methods in the paper. The empirical application is given in Section 6 and Section 7 concludes. Technical proofs are found in the appendix.

2 Model and assumptions

Although the results derived in this paper are of general applicability, it is helpful to discuss the model and derivations in light of the specific question of stock return predictability. Thus, let the dependent variable be denoted r_t , which would typically represent excess stock returns when analyzing return predictability, and the corresponding vector of regressors, x_t , where x_t is an $m \times 1$ vector and $t = 1, \dots, T$. The behavior of r_t and x_t are assumed to satisfy,

$$r_t = \alpha + \beta x_{t-1} + u_t, \tag{1}$$

$$x_t = Ax_{t-1} + v_t, \tag{2}$$

where $A = I + C/T$ is an $m \times m$ matrix. The error processes are assumed to satisfy the following conditions.

Assumption 1 *Let $w_t = (u_t, \epsilon_t)'$ and $\mathcal{F}_t = \{w_s | s \leq t\}$ be the filtration generated by w_t . Then*

1. $v_t = D(L)\epsilon_t = \sum_{j=0}^{\infty} D_j \epsilon_{t-j}$, and $\sum_{j=0}^{\infty} j \|D_j\| < \infty$.
2. $E[w_t | \mathcal{F}_{t-1}] = 0$.
3. $E[w_t w_t'] = \Sigma = [(\sigma_{11}, \sigma_{12}), (\sigma_{21}, I)]$.
4. $\sup_t E[u_t^4] < \infty$ and $\sup_t E[\|\epsilon_t\|^4] < \infty$.

The model described by equations (1) and (2) and Assumption 1 captures the essential features of a predictive regression with nearly persistent regressors. It states the usual martingale difference (mds) assumption for the errors in the return processes but allows for a linear time-series structure in the errors of the predictor variables; moreover, the innovations can be conditionally heteroskedastic, as long as they are covariance stationary. The error terms u_t and v_t are also often highly correlated. The auto-regressive roots of the regressors are parametrized as being local-to-unity, which captures the near-unit-root behavior of many predictor variables, but is less restrictive than a pure unit-root

assumption.

Similar models are used to analyze the predictability of stock returns by Cavanagh et al. (1995), Lanne (2002), Valkanov (2003), Torous et al. (2004), and Campbell and Yogo (2005).

Let $E_t = (u_t, v_t)'$ be the joint innovations process. Under Assumption 1, by standard arguments (Phillips and Solo, 1992),

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} E_t \Rightarrow B(r) = BM(\Omega)(r),$$

where $\Omega = [(\omega_{11}, \omega_{12}), (\omega_{21}, \Omega_{22})]$, $\omega_{11} = \sigma_{11}$, $\omega_{21} = D(1)\sigma_{12}$, $\omega_{12} = \omega'_{21}$, $\Omega_{22} = D(1)D(1)'$, and $B(\cdot) = (B_1(\cdot), B_2(\cdot))'$ denotes an $1 + m$ -dimensional Brownian motion. Also, let $\Lambda_{22} = \sum_{k=1}^{\infty} E(v_k v_0')$ be the one-sided long-run variance of v_t . The following lemma sums up the key asymptotic results for the nearly integrated model in this paper (Phillips 1987, 1988).

Lemma 1 *Under Assumption 1, as $T \rightarrow \infty$,*

- (a) $T^{-1/2} x_{i,[Tr]} \Rightarrow J_C(r)$,
 - (b) $T^{-3/2} \sum_{t=1}^T x_t \Rightarrow \int_0^1 J_C(r) dr$,
 - (c) $T^{-2} \sum_{t=1}^T x_t x_t' \Rightarrow \int_0^1 J_C(r) J_C(r)' dr$,
 - (d) $T^{-1} \sum_{t=1}^T u_t x_{t-1}' \Rightarrow \int_0^1 dB_1(r) J_C(r)'$,
 - (e) $T^{-1} \sum_{t=1}^T v_t x_{t-1}' \Rightarrow \int_0^1 dB_2(r) J_C(r)' + \Lambda_{22}$,
- where $J_C(r) = \int_0^r e^{(r-s)C} dB_2(s)$.

Analogous results hold for the demeaned variables $\underline{x}_t = x_t - T^{-1} \sum_{t=1}^n x_t$, with the limiting process J_C replaced by $\underline{J}_C = J_C - \int_0^1 J_C$. These results are used repeatedly below.

In both empirical work, and the applied theory literature, the stronger assumption that *both* u_t and v_t are martingale difference sequences is often imposed. In the current paper, I will primarily work with the general linear process structure for v_t , but I will also comment and derive some of the results under the simpler mds assumption.

Assumption 2 *Let $w_t = (u_t, v_t)'$ and $\mathcal{F}_t = \{w_s | s \leq t\}$ be the filtration generated by w_t . Then*

1. $E[w_t | \mathcal{F}_{t-1}] = 0$.
2. $E[w_t w_t'] = \Omega = [(\omega_{11}, \omega_{12}), (\omega_{21}, \Omega_{22})]$.
3. $\sup_t E[u_t^4] < \infty$ and $\sup_t E[\|v_t\|^4] < \infty$.

Before discussing the long-run estimators of β in equation (1) it is useful to quickly review the properties of the standard (short-run) OLS estimator of β , since the long-run estimators will reflect

many of these properties. Let $\hat{\beta}$ denote the standard OLS estimate of β in equation (1). By Lemma 1 and the continuous mapping theorem (CMT), it follows that

$$T(\hat{\beta} - \beta) \Rightarrow \left(\int_0^1 dB_1 J'_C \right) \left(\int_0^1 J_C J'_C \right)^{-1}, \quad (3)$$

as $T \rightarrow \infty$. This result is identical under either Assumption 1 or 2. Analogous to the case with pure unit-root regressors, the OLS estimator does not have an asymptotically mixed normal distribution due to the correlation between B_1 and B_2 , which causes B_1 and J_C to be correlated. Therefore, standard test procedures cannot be used.

The greatest problem in dealing with regressors that are near-unit-root processes is the nuisance parameter C . If C was known, critical values for the asymptotic distribution in (3) could be obtained from simulations, for instance. More efficient methods can also be implemented in this case as discussed below in Section 3.3. However, C is generally unknown and not consistently estimable. It is nevertheless useful to first derive inferential methods under the assumption that C is known, and then use the arguments of Cavanagh et al. (1995) to construct feasible tests. The following section derives and outlines the inferential methods used for estimating and performing tests on β in equation (1), treating C as known. Section 5 discusses how the methods of Cavanagh et al. (1995), and Campbell and Yogo (2005), can be used to construct feasible tests with C unknown.

3 Long-run estimation

3.1 The fitted regressions

In long-run regressions, the focus of interest are fitted regressions of the type

$$r_{t+q}(q) = \alpha^U(q) + \beta^U(q) x_t + u_{t+q}(q), \quad (4)$$

and

$$r_{t+q}(q) = \alpha^B(q) + \beta^B(q) x_t(q) + u_{t+q}(q), \quad (5)$$

where $r_t(q) = \sum_{j=1}^q r_{t-q+j}$ and $x_t(q) = \sum_{j=1}^q x_{t-q+j}$. In equation (4), long-run future returns are regressed onto a one period predictor, whereas in equation (5), long-run future returns are regressed

onto long-run past regressors. Equation (4) is the specification most often used for testing stock return predictability, although Fama and French (1988b) use (5) in a univariate framework where sums of future returns are regressed onto sums of past returns. For completeness, I will consider both kinds of specifications here. The regressions in equation (4) and (5) will be referred to as the unbalanced and balanced regressions, respectively, since in the former case long-run returns are regressed onto short-run predictors and in the latter long-run returns are regressed onto long-run predictors. This choice of terminology, i.e. unbalanced and balanced, is used purely as a mnemonic device; ‘unbalanced’ is not meant to convey anything negative about this specification.

Let the OLS estimators of $\beta^U(q)$ and $\beta^B(q)$ in equations (4) and (5), using overlapping observations, be denoted by $\hat{\beta}^U(q)$ and $\hat{\beta}^B(q)$, respectively. A long-standing issue is the calculation of correct standard errors for $\hat{\beta}^U(q)$ and $\hat{\beta}^B(q)$. Since overlapping observations are used to form the estimates, the residuals $u_t(q)$ will exhibit serial correlation; standard errors failing to account for this fact will lead to biased inference. The common solution to this problem has been to calculate auto-correlation robust standard errors, using methods described by Hansen and Hodrick (1980) and Newey and West (1987). However, these robust estimators tend to have rather poor finite sample properties; this is especially so in cases when the serial correlation is strong, as it often is when overlapping observations are used. In this section, I derive the asymptotic properties of $\hat{\beta}^U(q)$ and $\hat{\beta}^B(q)$ under the assumption that the forecasting horizon q grows with the sample size but at a slower pace. The results complement those of Valkanov (2003), who treats the case where the forecasting horizon grows at the same rate as the sample size. Simulation results in Valkanov (2003) and this paper show that both asymptotic approaches provide limiting distributions that are good proxies for the finite sample behaviour of the long-run estimators. However, it turns out that the asymptotic results derived under the current assumptions are more transparent than those of Valkanov (2003) and provide further understanding of the properties of the long-run estimators, both under the null and the alternative hypothesis of predictability. In particular, the results here show the strong connection between the limiting distributions of the short- and long-run estimators. This finding has important implications for the construction of more efficient estimators and test-statistics that control for the endogeneity and persistence in the regressors. The procedures in this paper also avoid the need for simulation methods; the proposed test-statistics have limiting normal distributions, although in the case of endogenous regressors with unknown persistence, Bonferroni type methods need to be used to construct tests and confidence intervals.

3.2 The limiting distributions of the long-run OLS estimators

Given that equations (4) and (5) are estimated with overlapping observations, created from short-run data, they should be viewed as fitted regressions rather than actual data generating processes (dgp); the use of overlapping observations effectively necessitates the specification of a dgp for the observed short-run data. The results below are derived under the assumption that the true dgp satisfies equations (1) and (2), and that the long-run observations are formed by summing up data generated by that process. Under the null hypothesis of no predictability, the one period dgp is simply $r_t = u_t$, in which case the long-run coefficients $\beta^U(q)$ and $\beta^B(q)$ will also be equal to zero. It follows that under the null both equations (4) and (5) are correctly specified and the analysis of $\hat{\beta}^U(q)$ and $\hat{\beta}^B(q)$ simplifies. It is therefore common in the literature to only derive asymptotic results for long-run estimators under the null of no predictability. By considering the properties of the estimators both under the null and the alternative, however, a more complete picture of the properties of the long-run estimators emerges. Of course, equation (1) is only one possible alternative to the null of no predictability, but it provides a benchmark case.

Theorem 1 *Suppose the data is generated by equations (1) and (2), and that Assumption 1 holds.*

1. *Under the null hypothesis that $\beta = 0$, as $q, T \rightarrow \infty$, such that $q/T \rightarrow 0$,*

(a)

$$\frac{T}{q} \left(\hat{\beta}^U(q) - 0 \right) \Rightarrow \left(\int_0^1 dB_1 \underline{J}'_C \right) \left(\int_0^1 \underline{J}_C \underline{J}'_C \right)^{-1}, \quad (6)$$

(b)

$$T \left(\hat{\beta}^B(q) - 0 \right) \Rightarrow \left(\int_0^1 dB_1 \underline{J}'_C \right) \left(\int_0^1 \underline{J}_C \underline{J}'_C \right)^{-1}. \quad (7)$$

2. *Under the alternative hypothesis that $\beta \neq 0$, as $q, T \rightarrow \infty$, such that $q/T \rightarrow 0$,*

(a)

$$\frac{2T}{q^2} \left(\hat{\beta}^U(q) - \beta^U(q) \right) \Rightarrow \beta \left(\int_0^1 dB_2 \underline{J}'_C + \Lambda_{22} \right) \left(\int_0^1 \underline{J}_C \underline{J}'_C \right)^{-1}, \quad (8)$$

(b)

$$\frac{T}{q} \left(\hat{\beta}^B(q) - \beta^B(q) \right) \Rightarrow \beta \left(\int_0^1 dB_2 \underline{J}'_C + \Omega_{22} \right) \left(\int_0^1 \underline{J}_C \underline{J}'_C \right)^{-1}, \quad (9)$$

where $\beta^U(q) = \beta (I + A + \dots + A^{q-1})$ and $\beta^B(q) = \beta A^{q-1}$. Since $A = I + C/T$, it follows that $\beta^U(q)/q = \beta + O(q/T) \rightarrow \beta$ and $\beta^B(q) = \beta + O(q/T) \rightarrow \beta$, as $q, T \rightarrow \infty$, such that $q/T \rightarrow 0$.

Theorem 1 shows that under the null of no predictability, the limiting distributions of $\hat{\beta}^U(q)$ and $\hat{\beta}^B(q)$ are identical to that of the plain short-run OLS estimator $\hat{\beta}$, although $\hat{\beta}^U(q)$ needs to be standardized by q^{-1} , since, as seen in part 2 of the theorem, the estimated parameter $\beta^U(q)$ is of an order q times larger than the original short-run parameter β .

Under the alternative hypothesis of predictability, the limiting distributions of $\hat{\beta}^U(q)$ and $\hat{\beta}^B(q)$ are quite different from the short-run result, and are in fact similar to the distribution of the OLS estimator of the first order auto-regressive root in x_t , although the rate of convergence is slower. The estimators still converge to well defined parameters under the alternative hypothesis, but their asymptotic distributions are driven by the auto-regressive nature of the regressors and the fact that the fitted regressions in (4) and (5) are effectively misspecified, under the assumption that the true relationship takes the form of equation (1). In addition, the limiting distributions of $\hat{\beta}^U(q)$ and $\hat{\beta}^B(q)$ exhibit second order bias terms in the shape of Λ_{22} and Ω_{22} , respectively. These bias terms do not show up in the asymptotics of the short-run ($q = 1$) OLS estimator, and are, again, a result of the effective misspecification of the long-run regression. Interestingly, under Assumption 2, the second order bias term in $\hat{\beta}^U(q)$ disappears, since in this case $\Lambda_{22} = 0$, while the second order bias in $\hat{\beta}^B(q)$ still remains.

The equality between the long-run asymptotic distributions under the null hypothesis, shown in Theorem 1, and that of the short-run OLS estimator in equation (3), may seem puzzling. To understand the intuition behind this result, it is helpful to consider the asymptotic properties of narrow band frequency domain regressions around the zero frequency. Since such regressions also attempt to capture a long-run relationship, they are similar in nature to the long-run regressions considered here. As shown by Phillips (1991b) and Corbae et al. (2002), when the regressors are unit-root processes, the limiting distribution of the (non-parametric) narrow band zero frequency estimator is identical to that of the standard OLS time-domain estimator, or equivalently to the full frequency band spectral regression. There is thus no asymptotic efficiency loss from excluding all frequencies outside a neighbourhood around the origin. The reason behind this result is that the spectra of the unit-root processes is dominated by the behaviour at the zero frequency. That is, the long-run movements in the regressors dominate their behaviour, and ignoring the short-run movements causes no asymptotic efficiency loss. The same phenomenon occurs for the long-run estimators considered here, which explains the identical rates of convergence and limiting distributions for the short- and long-run estimators.

3.3 Endogeneity corrections

It is apparent that, under the null hypothesis, the long-run OLS estimators suffer from the same endogeneity problems as the short-run estimator. It is therefore useful to first consider suitable endogeneity corrections in the short-run regression; the following discussion repeats the results in Hjalmarrsson (2005).

In the pure unit-root case, one popular inferential approach is to “fully modify” the OLS estimator as suggested by Phillips and Hansen (1990) and Phillips (1995). This estimator is suitable under Assumption 1, since it allows for a very general error structure. In the near-unit-root case, a similar method can be considered. Define the quasi-differencing operator

$$\Delta_C x_t = x_t - x_{t-1} - \frac{C}{T} x_{t-1} = v_t, \quad (10)$$

and let $r_t^+ = r_t - \hat{\omega}_{12} \hat{\Omega}_{22}^{-1} \Delta_C x_t$ and $\hat{\Lambda}_{12}^+ = -\hat{\omega}_{12} \hat{\Omega}_{22}^{-1} \hat{\Lambda}_{22}$, where $\hat{\omega}_{12}$, $\hat{\Omega}_{22}^{-1}$, and $\hat{\Lambda}_{22}$ are consistent estimates of the respective parameters.³ The fully modified OLS estimator is now given by

$$\hat{\beta}_{FM}^+ = \left(\sum_{t=1}^T r_t^+ \underline{x}'_{t-1} - T \hat{\Lambda}_{12}^+ \right) \left(\sum_{t=1}^T \underline{x}_{t-1} \underline{x}'_{t-1} \right)^{-1}, \quad (11)$$

where $\underline{r}_t^+ = \underline{r}_t - \hat{\omega}_{12} \hat{\Omega}_{22}^{-1} \Delta_C x_t$ and $\underline{r}_t = r_t - T^{-1} \sum_{t=1}^t r_t$. The only difference in the definition of (11), from the FM-OLS estimator for the pure unit-root case, is the use of the quasi-differencing operator, as opposed to the standard differencing operator. Once the innovations v_t are obtained from quasi-differencing, the modification proceeds in exactly the same manner as in the unit-root case.

Define $\omega_{11.2} = \omega_{11} - \omega_{12} \Omega_{22}^{-1} \omega_{21}$ and the Brownian motion $B_{1.2} = B_1 - \omega_{12} \Omega_{22}^{-1} B_2 = BM(\omega_{11.2})$. The process $B_{1.2}$ is now orthogonal to B_2 and J_C . Using the same arguments as Phillips (1995), it follows that, as $T \rightarrow \infty$,

$$T \left(\hat{\beta}_{FM}^+ - \beta \right) \Rightarrow \left(\int_0^1 dB_{1.2} J_C' \right) \left(\int_0^1 J_C J_C' \right)^{-1} \equiv MN \left(0, \omega_{11.2} \left(\int_0^1 J_C J_C' \right)^{-1} \right). \quad (12)$$

The corresponding t -statistics and Wald statistics will have standard distributions asymptotically.

Campbell and Yogo (2005) derive similar results, but they attack the problem from a test point-of-view, whereas the derivation by Hjalmarrsson (2005) that is presented here starts with the estimation

³The definition of $\hat{\Lambda}_{12}^+$ is slightly different from the one found in Phillips (1995). This is due to the predictive nature of the regression equation (1), and the martingale difference sequence assumption on u_t .

problem. However, presenting the derivation in this manner makes clear that this approach is a generalization of fully modified estimation.

Analogously, a generalization of the augmented regression equation of Phillips (1991a) is possible. By obtaining $v_t = \Delta_C x_t$ through quasi-differencing, the augmented regression

$$r_t = \alpha + \beta x_{t-1} + \gamma v_t + u_{t,2} \quad (13)$$

can be estimated. Under Assumption 2, the OLS estimator of β in equation (13) will have an asymptotic distribution identical to that in (12).

Similar remedies to those just presented for the short-run case, such as the fully modified approach, can be considered in the long-run. However, simulations not reported in the paper show that the finite sample properties of the fully-modified long-run estimators appear very unsatisfactory. In the sequel I therefore focus on a long-run version of the augmented regression equation in (13), which turn out to work well in finite samples. The asymptotic results for the long-run FM-OLS estimator can be derived in a similar manner and are not given here.

Consider the fitted augmented regression equations

$$r_{t+q}(q) = \alpha^U(q) + \beta^U(q) x_t + \gamma^U(q) v_{t+q}(q) + u_{t+q,2}(q), \quad (14)$$

and

$$r_{t+q}(q) = \alpha^B(q) + \beta^B(q) x_t(q) + \gamma^B(q) v_{t+q}(q) + u_{t+q,2}(q), \quad (15)$$

where $v_t(q) = \sum_{j=1}^q v_{t-q+j}$. Let $\hat{\beta}^{U+}(q)$ and $\hat{\beta}^{B+}(q)$ be the OLS estimators of $\beta^U(q)$ and $\beta^B(q)$ in equations (14) and (15).

Theorem 2 *Suppose the data is generated by equations (1) and (2), and that Assumption 2 holds.*

1. *Under the null hypothesis that $\beta = 0$, as $q, T \rightarrow \infty$, such that $q/\sqrt{T} \rightarrow 0$,*

$$T \left(\hat{\beta}^{B+}(q) - 0 \right), \frac{T}{q} \left(\hat{\beta}^{U+}(q) - 0 \right) \Rightarrow MN \left(0, \omega_{11,2} \left(\int_0^1 \underline{J}_C \underline{J}'_C \right)^{-1} \right), \quad (16)$$

where $\omega_{11,2} = \omega_{11} - \omega_{12} \Omega_{22}^{-1} \omega_{21}$.

2. Under the alternative hypothesis that $\beta \neq 0$, as $q, T \rightarrow \infty$, such that $q/\sqrt{T} \rightarrow 0$,

$$\frac{T}{q} \left(\hat{\beta}^{B+}(q) - \beta^B(q) \right), \frac{2T}{q^2} \left(\hat{\beta}^{U+}(q) - \beta^U(q) \right) \Rightarrow \beta \left(\int_0^1 dB_2 \underline{J}'_C \right) \left(\int_0^1 \underline{J}_C \underline{J}'_C \right)^{-1}. \quad (17)$$

Under the null hypothesis of no predictability, the estimators $\hat{\beta}^{U+}(q)$ and $\hat{\beta}^{B+}(q)$ have asymptotically mixed normal distributions, although under the alternative hypothesis of predictability, the asymptotic distributions are still non-standard. It is interesting to note, however, that the second order bias term that appears in the balanced long-run OLS estimator $\hat{\beta}^B(q)$, under the alternative hypothesis, is in fact eliminated by fitting the augmented regression equation.

The results in Theorems 1 and 2 bring some clarity to the properties of long-run regressions with nearly persistent regressors. Under the null of no predictability, the long-run estimators have identical asymptotic distributions to the short-run estimators. Under the alternative hypothesis of predictability, however, the asymptotic properties of the long-run estimators change substantially and the results are now driven by the *de facto* misspecification of the long-run regressions, and the auto-regressive nature of the regressors; this is manifest in both the slower rate of convergence as well as the non-standard limiting distribution.

3.4 A normally distributed t -statistic for long-run inference

Given the asymptotically mixed normal distributions of $\hat{\beta}^{U+}(q)$ and $\hat{\beta}^{B+}(q)$ under the null hypothesis, standard test procedures can now be applied to test the null of no predictability. In fact, the following convenient result is easy to prove.

Corollary 1 *Let $t^{U+}(q)$ and $t^{B+}(q)$ denote the standard t -statistics corresponding to $\hat{\beta}^{U+}(q)$ and $\hat{\beta}^{B+}(q)$. That is,*

$$t^{U+}(q) = \frac{\hat{\beta}_k^{U+}(q)}{\sqrt{\left(\frac{1}{T} \sum_{t=1}^T \hat{u}^{U+}(q)^2 \right) a' \left(\sum_{t=1}^T \underline{z}_t \underline{z}'_t \right)^{-1} a}}, \quad (18)$$

and

$$t^{B+}(q) = \frac{\hat{\beta}_k^{B+}(q)}{\sqrt{\left(\frac{1}{T} \sum_{t=1}^T \hat{u}^{B+}(q)^2 \right) a' \left(\sum_{t=1}^T \underline{z}_t(q) \underline{z}'_t(q) \right)^{-1} a}}, \quad (19)$$

where $\hat{u}^{U+}(q)$ and $\hat{u}^{B+}(q)$ are the estimated residuals, $\underline{z}_t = (\underline{x}_t, v_{t+q}(q))$, $\underline{z}_t(q) = (\underline{x}_t(q), v_{t+q}(q))$ and a is an $2m \times 1$ vector with the k 'th component equal to one and zero elsewhere. Then, under

Assumption 2 and the null-hypothesis of $\beta = 0$, as $q, T \rightarrow \infty$, such that $q/\sqrt{T} \rightarrow 0$,

$$\frac{t^{U+}(q)}{\sqrt{q}}, \frac{t^{B+}(q)}{\sqrt{q}} \Rightarrow N(0, 1). \quad (20)$$

Thus, long-run inference can be performed by simply scaling the corresponding standard t -statistic by $q^{-1/2}$. In the case of exogenous regressors, such that $\omega_{12} = 0$, the scaling result in (20) will of course apply to the t -statistics from the standard long-run estimators $\hat{\beta}^U(q)$ and $\hat{\beta}^B(q)$, which makes long-run inference extremely straightforward in this case.

4 Monte Carlo results

All of the above asymptotic results are derived under the assumption that the forecasting horizon grows with the sample size, but at a slower rate. Valkanov (2003) and Torous et al. (2004) also study long-run regressions with near-integrated regressors, but derive their asymptotic results under the assumption that $q/T \rightarrow \kappa \in (0, 1)$ as $q, T \rightarrow \infty$. That is, they assume that the forecasting horizon grows at the same pace as the sample size. Under such conditions, the asymptotic properties of $\hat{\beta}^U(q)$ and $\hat{\beta}^B(q)$ are quite different from those derived in this paper. There is, of course, no right or wrong way to perform the asymptotic analysis; what matters in the end is how well the asymptotic distributions capture the properties of actual finite sample estimates. To this end, a brief Monte Carlo simulation is therefore conducted.

Equations (1) and (2) are simulated, with u_t and v_t drawn from an *iid* bivariate normal distribution with mean zero, unit variance and correlation $\delta = -0.9$. The large negative correlation is chosen to assess the effectiveness of the endogeneity corrections in $\hat{\beta}^{U+}(q)$ and $\hat{\beta}^{B+}(q)$, as well as to reflect the sometimes high endogeneity of regressors such as the dividend- or earnings-price ratio in tests of stock return predictability. The intercept α is set to one and the local-to-unity parameter C is set to either 0 or -10 . Three different estimators, and their corresponding t -statistics, are considered: the long-run estimators, $\hat{\beta}^{U+}(q)$ and $\hat{\beta}^{B+}(q)$, as well as the short-run OLS estimator in the augmented regression equation (13).⁴ Since the aim of the simulation is to determine how well the asymptotic distributions derived above reflect actual finite sample distributions, all estimation and testing is done under the assumption that the parameter C is known. The sample sizes are chosen as $T = 100$ and $T = 500$.

⁴As shown by Phillips (1991a), in the case of normally distributed errors, the OLS estimator in the short-run augmented regression equation (13) will in fact be equal to the maximum likelihood estimator.

The first part of the simulation study evaluates the finite sample properties of the three estimators under an alternative of predictability, where the true β is set equal to 0.05 and the second part analyzes the size and power properties of the scaled t -tests. The forecasting horizon is set to $q = 12$ and $q = 60$ for the $T = 100$ and $T = 500$ samples, respectively. These forecasting horizons are similar to those often used in practice for similar sample sizes. All results are based on 100,000 repetitions.

The results are shown in Figures 1 and 2 for $C = 0$ and -10 , respectively. In the top two panels, (A1) and (A2) in both figures, the kernel estimates of the densities of the estimated coefficients are shown. To enable a fair comparison, $\hat{\beta}^{U+}(q)$ is divided by $(1 + A + \dots + A^{q-1})$ and $\hat{\beta}^{B+}(q)$ by A^{q-1} , as indicated by the result in Theorem 1. The non-standard distributions of $\hat{\beta}^{U+}(q)$ and $\hat{\beta}^{B+}(q)$ under the alternative are evident, especially so for $\hat{\beta}^{B+}(q)$. The fact that $\hat{\beta}^{U+}(q)$ converges faster than $\hat{\beta}^{B+}(q)$ under the alternative, after proper scaling, is also clear. This is especially clear in the $C = -10$ case shown in Figure 2. The short-run estimator outperforms both long-run estimators, however. In the bottom graphs, (B1) and (B2), the rejection rates of the 5% two-sided t -tests, for tests of the null of no predictability, are given. For both $T = 100$ and $T = 500$, all three tests have a rejection rate very close to 5% under the null, so the scaling of the long-run t -statistics by $q^{-1/2}$ appears to work well in practice, as well as the endogeneity corrections. For $C = -10$, there are some slight deviations from the nominal 5% size in the average rejection rates under the null. These deviations are small, however, and well within the acceptable range; the rejection rates for $t^{U+}(q)/\sqrt{q}$ and $t^{B+}(q)/\sqrt{q}$ under the null of no predictability are 0.041 and 0.077, respectively, for $T = 100$, and 0.041 and 0.074 for $T = 500$. The test based on the $\hat{\beta}^{U+}(q)$ estimator has similar power properties to the short-run test, although the short-run test performs better in all instances. The test based on $\hat{\beta}^{B+}(q)$ performs rather poorly in all cases.

In summary, the simulation results show that the endogeneity correction performed in $\hat{\beta}^{U+}(q)$ and $\hat{\beta}^{B+}(q)$ appears to work well and that the scaling of the t -statistic, as suggested by Corollary 1, achieves the correct size. Both the asymptotically slower rate of convergence for $\hat{\beta}^{U+}(q)$ and $\hat{\beta}^{B+}(q)$ under the alternative of predictability and the finite sample results given in Figures 1 and 2 indicate that there is little reason to consider long-run tests if one believes that the alternative model of stock return predictability is given by equation (1).

Simulations not reported in this paper also show that these results are robust to heteroskedastic error terms in the form of ARCH/GARCH processes.

5 Feasible methods

To implement the methods described in the two previous sections, knowledge of the parameter C is required. Since C is typically unknown and not estimable in general, the bounds procedures of Cavanagh et al. (1995) and Campbell and Yogo (2005) can be used to obtain feasible procedures. The following discussion assumes a scalar regressor, as do the above studies.

Although C is not estimable, a confidence interval for C can be obtained, as described by Stock (1991). By evaluating the estimator and test-statistic for each value of C in that confidence interval, a range of possible estimates and values of the test-statistic are obtained. A conservative test can then be formed by choosing the most conservative value of the test statistic, given the alternative hypothesis. If the confidence interval has a coverage rate of $100(1 - \alpha_1)\%$ and the nominal size of the test is α_2 , then by Bonferroni's inequality the final conservative test will have a size no greater than $\alpha = \alpha_1 + \alpha_2$. In general, the size of the test will be less than α , and a test with a pre-specified size can be achieved by fixing α_2 and adjusting α_1 . Since a test-statistic will by necessity be conservative only against a one-sided alternative, conservative confidence intervals provide a more convenient way of presenting the results from the Bonferroni tests. Let \overline{C} and \underline{C} denote the values in the $100(1 - \alpha_1)\%$ confidence interval for C , for which the test-statistic takes on the largest and smallest value, respectively. The lower bound of the $100(1 - \alpha_1 - \alpha_2)\%$ confidence interval for $\beta^U(q)$ is then given by $\hat{\beta}_{\overline{C}}^{U+}(q) - z_{\alpha_2/2} \hat{\beta}_{\overline{C}}^{U+}(q) \sqrt{q}/t_{\overline{C}}^{U+}(q)$ and the upper bound by $\hat{\beta}_{\underline{C}}^{U+}(q) + z_{\alpha_2/2} \hat{\beta}_{\underline{C}}^{U+}(q) \sqrt{q}/t_{\underline{C}}^{U+}(q)$, where the \overline{C} and \underline{C} subscripts indicate that the estimators and the test-statistics are evaluated for these values of C and $z_{\alpha_2/2}$ denotes the $1 - \alpha_2/2$ quantile of the standard normal distribution. Again, α_1 can be adjusted to achieve a desired $100(1 - \alpha)\%$ coverage rate. Analogously, a confidence interval for $\beta^B(q)$ can be constructed from the estimator $\hat{\beta}^{B+}(q)$. Note that, unlike the short-run methods in Campbell and Yogo (2005), there is no guarantee that \overline{C} and \underline{C} are the endpoints of the confidence interval for C , although for most values of q they typically are; the test-statistics should thus be evaluated for all values in C , to find \overline{C} and \underline{C} .

In the empirical section, I adopt a similar approach to Campbell and Yogo (2005). A confidence interval for C is obtained by inverting the DF-GLS unit-root statistic. Table 2 of Campbell and Yogo (2005) is used to find the desired significance level of this confidence interval in order for the confidence interval for the long-run slope coefficient to have a 90% coverage rate. A drawback of this method is that no clear-cut point estimate is produced, but rather a range of estimates. I therefore present

standard long-run OLS point estimates in the empirical work below.

6 Empirical application to stock return predictability

To illustrate the theoretical results derived in this paper, I revisit the question of stock return predictability. There have been many conflicting results regarding the existence of a predictable component in stock returns. However, the most recent work by Lewellen (2004) and Campbell and Yogo (2005), which rely on both robust as well as more efficient methods of inference, do find evidence that stock returns are predictable to some degree. In this section, I extend their empirical analysis to the long-horizon case. Since the long-run estimators $\hat{\beta}^{U+}(q)$ and $\hat{\beta}^{B+}(q)$, which control for the endogeneity and persistence in the regressors, and the corresponding scaled t -tests are effectively long-run versions of the methods developed in Campbell and Yogo (2005), the empirical results presented here provide a direct comparison with previous work.

To further facilitate comparison between the short-run and the long-run case, I use a subset of the data used by Campbell and Yogo (2005); these data were downloaded from Professor Yogo's website.⁵ They consist of the annual excess returns on the S&P 500 over the period 1880-2002 as well as the monthly excess returns on the CRSP NYSE/AMEX value-weighted index over the period 1926-2002. The excess returns are calculated as the stock returns over the riskfree rate, measured by the return on the one-month T-bill for the monthly data, and by the return on the three-month T-bill rolled over quarterly for the annual data. As predictor variables, I use the dividend-price ratio and the smoothed earnings-price ratio suggested by Campbell and Shiller (1988). The dividend price ratio is calculated as dividends over the past year divided by the current price and the (smoothed) earnings-price ratio as the average earnings of the past 10 years divided by the current price. Since earnings are not available for the CRSP data, the corresponding S&P 500 earnings are used for this series as well. All regressions are run using log-transformed variables with the log excess returns as the dependent variable.

The two key data characteristics that define the properties of the regression estimators analyzed in this paper are the near persistence and endogeneity of the regressors. Table 1 shows confidence intervals for the local-to-unity parameter C , calculated by inverting the DF-GLS unit-root test, as well as estimates of the correlation between the innovations to returns and the innovations to the regressors; these estimates are labeled δ in Table 1. The values are calculated in an identical manner to those of

⁵<http://finance.wharton.upenn.edu/~yogo/>

Table 4 in Campbell and Yogo (2005). As is evident, there is a large negative correlation between the innovations to returns and the predictor variables, both for the dividend- and earnings-price ratios. With the exception of the annual earnings-price ratio series, the confidence intervals for C include the unit-root case and it is clear that all predictor variables have first order auto-regressive roots that are close to unity. There is thus a strong case for using estimators that correct for the bias induced by the endogeneity and persistence in the regressors.

Given the poor performance of the $\hat{\beta}^{B+}(q)$ estimator in the simulation exercises, I only show results for the estimator $\hat{\beta}^{U+}(q)$; the standard version, $\hat{\beta}^U(q)$, of this estimator is also what is typically used in long-run tests of stock return predictability. To illustrate the effects of an increasing forecasting horizon, I show the estimates, with confidence intervals, as plots against the forecasting horizon q . Figures 3 and 4 show the results for the annual S&P 500 returns and the monthly CRSP returns, respectively. The left-hand plots in these figures show the 90% Bonferroni confidence intervals based on the estimator $\hat{\beta}^{U+}(q)$ and the right hand plots show the standard long-run OLS point estimate $\hat{\beta}^U(q)$, along with 90% confidence intervals. The OLS confidence intervals, calculated as $\hat{\beta}^U(q) \pm 1.65 \times \hat{\beta}^U(q) \sqrt{q}/t^U(q)$, are shown as a comparison, but are not valid since the long-run OLS estimator is not mixed normally distributed; that is, these confidence intervals take into account the overlap in the data, but not the endogeneity and persistence in the regressors. The Bonferroni confidence intervals are calculated according to the methods described in the previous section, based on an inversion of the DF-GLS statistic. The estimates and confidence intervals have all been scaled by q^{-1} , as suggested by Theorem 1, to ease comparison across different horizons.

At the one-period horizon, the results are qualitatively identical to those of Campbell and Yogo;⁶ based on the Bonferroni confidence intervals shown in the right hand side graphs of Figures 3 and 4, there is evidence that the earnings-price ratio predicts stock-returns both in the annual and the monthly data, but the null of no predictability cannot be rejected when using the dividend price ratio as a regressor. If standard OLS methods were used to make inference, the null hypothesis would be rejected in all cases, for both the earnings- and dividend-price ratio, at the one-period horizon. What is interesting, however, is the behaviour of the confidence intervals as the forecasting horizon increases. Contrary to some popular beliefs, the evidence in favour of predictability is not stronger at the longer horizons. In fact, the case for predictability typically becomes weaker as the forecasting

⁶They are not numerically identical since the confidence intervals based on $\hat{\beta}^{U+}(q)$, with $q = 1$, will be slightly different from those based on Campbell and Yogo's Q -test. Also, unlike here, all results in Campbell and Yogo are standardized to conform with a model where both innovation processes have unit variance.

horizon increases. Only in the monthly data, with the dividend-price ratio as a regressor, is there slightly stronger evidence at longer horizons. However, the long-run evidence is extremely weak and seems more like random fluctuations than a consistent pattern; the confidence intervals are pointwise along q and the overall confidence level in the monthly case could thus be substantially less than the nominal pointwise level. It is also clear that failure to control for the endogeneity and persistence in the regressors will vastly overstate the case for predictability, in the long-run as well as the short-run. This is particularly evident in the results for the monthly data, shown in Figure 4.

Given the theoretical results, and the simulation evidence presented earlier in this paper, these empirical results are not hard to reconcile. Since the evidence of predictability is weak also at the short horizon, and there is a demonstrated loss of power as the forecasting horizon increases, it is not surprising that the null of no predictability cannot be rejected for longer horizons.

These results also generally agree with those of Ang and Bekaert (2003) who also draw the conclusion that the evidence of predictability tends to be strongest at a shorter horizon. Their findings suggest that previous strong evidence of long-run predictability may be largely due to incorrect inference based on Hansen and Hodrick (1980) and Newey and West (1987) methods, as well as a much stronger case for predictability in data ending before 1990. Ang and Bekaert (2003) use Hodrick (1992) auto-correlation robust standard errors, which they argue have good finite sample properties. However, these rely on the regressors being covariance stationary, which is a restrictive assumption for variables like the dividend- and earnings-price ratio, as evidenced by the results in Table 1. Boudoukh et al. (2005) explicitly question the prevailing view of long-horizon predictability and reach similar conclusions to those presented here, although their focus is on the joint properties of the regression estimators across different horizons.

7 Conclusion

I derive several new results for long-horizon regressions that use overlapping observations when the regressors are endogenous and nearly persistent. I show how to properly correct for the overlap in the data in a simple manner that obviates the need for auto-correlation robust standard error methods in these regressions. Further, when the regressors are persistent and endogenous, I show how to correct the long-run OLS estimators and test procedures in a manner similar to that proposed by Campbell and Yogo (2005) for the short-run case.

In addition, by analyzing the asymptotic properties of the long-run estimators under an alternative of predictability, I find strong evidence that long-horizon tests tend to have less power to detect deviations from the null hypothesis of no predictability than do one-period short-run tests.

An empirical application to stock-return predictability illustrates these results and shows that, in line with the theoretical results of this paper, the evidence for predictability is typically weaker as the forecasting horizon gets longer, reflecting the loss of power in long-run tests.

A Proofs

Proof of Theorem 1. For ease of notation the case with no intercept is treated. The results generalize immediately to regressions with fitted intercepts by replacing all variables by their demeaned versions. All limits as $q, T \rightarrow \infty$ are under the condition that $q/T \rightarrow 0$.

1. (a) Under the null hypothesis,

$$\frac{T}{q} \left(\hat{\beta}^U(q) - 0 \right) = \left(\frac{1}{qT} \sum_{t=1}^T u_{t+q}(q) x'_t \right) \left(\frac{1}{T^2} \sum_{t=1}^T x_t x'_t \right)^{-1} = \left(\frac{1}{qT} \sum_{t=1}^T \sum_{j=1}^q u_{t+j} x'_t \right) \left(\frac{1}{T^2} \sum_{t=1}^T x_t x'_t \right)^{-1}.$$

By standard arguments,

$$\frac{1}{qT} \sum_{t=1}^T \sum_{j=1}^q u_{t+j} x'_t = \frac{1}{qT} \sum_{t=1}^T (u_{t+1} x'_t + \dots + u_{t+q} x'_t) \Rightarrow \int_0^1 dB_1 J'_C,$$

as $q, T \rightarrow \infty$, such that $q/T \rightarrow 0$, since for any $h > 0$, $\frac{1}{T} \sum_{t=1}^T u_{t+h} x'_t \Rightarrow \int_0^1 dB_1 J'_C$. Therefore,

$$\frac{T}{q} \left(\hat{\beta}^U(q) - 0 \right) \Rightarrow \left(\int_0^1 dB_1 J'_C \right) \left(\int_0^1 J_C J'_C \right)^{-1}.$$

(b) Under the null hypothesis,

$$\begin{aligned} T \left(\hat{\beta}^B(q) - 0 \right) &= \left(\frac{1}{q^2 T} \sum_{t=1}^T u_{t+q}(q) x_t(q) (q)' \right) \left(\frac{1}{q^2 T^2} \sum_{t=1}^T x_t(q) x_t(q)' \right)^{-1} \\ &= \left(\frac{1}{q^2 T} \sum_{t=1}^T \sum_{j=1}^q \sum_{k=1}^q u_{t+j} x'_{t-q+k} \right) \left(\frac{1}{q^2 T^2} \sum_{t=1}^T \sum_{j=1}^q \sum_{k=1}^q x_{t-q+j} x'_{t-q+k} \right)^{-1}. \end{aligned}$$

By some algebraic manipulations,

$$\begin{aligned}
& \frac{1}{T^2} \sum_{t=1}^T \sum_{j=1}^q \sum_{k=1}^q x_{t-q+j} x'_{t-q+k} \\
= & \hat{\gamma}_{xx}^1(0) + \hat{\gamma}_{xx}^2(0) + \dots + \hat{\gamma}_{xx}^{q-1}(0) + \hat{\gamma}_{xx}^q(0) + \hat{\gamma}_{xx}^2(-1) + \hat{\gamma}_{xx}^3(-1) + \dots + \hat{\gamma}_{xx}^q(-1) \\
& \vdots \\
& + \hat{\gamma}_{xx}^q(-(q-1)) + \hat{\gamma}_{xx}^1(q-1) \\
& \vdots \\
& + \hat{\gamma}_{xx}^{q-1}(1) + \hat{\gamma}_{xx}^{q-2}(1) + \dots + \hat{\gamma}_{xx}^1(1),
\end{aligned}$$

where $\hat{\gamma}_{xx}^k(h) = \frac{1}{T^2} \sum_{t=q}^{T-q+1} x_{t-q+k} x'_{t-q+k+h}$. Further, define $\gamma_{xx}(h) = \int_0^1 J_C J'_C$, and $\hat{\gamma}_{xx}(h) = \hat{\gamma}_{xx}^q(h) = \frac{1}{T^2} \sum_{t=q}^{T-q+1} x_t x'_{t+h}$. By standard arguments, for any fixed k and h ,

$$\hat{\gamma}_{xx}^k(h) \Rightarrow \gamma_{xx}(h) = \int_0^1 J_C J'_C,$$

since $q/T = o(1)$. Now, using Skorohod's representation theorem, there exists a probability space with random variables $\{\hat{\gamma}_{xx}^{k*}(h), \gamma_{xx}^*(h)\}$, for which $\hat{\gamma}_{xx}^{k*}(h) \xrightarrow{a.s.} \gamma_{xx}^*(h)$, $\hat{\gamma}_{xx}^{k*}(h) \equiv \hat{\gamma}_{xx}^k(h)$, and $\int_0^1 J_C^* J_C^{*'} \equiv \int_0^1 J_C J'_C$, where ' \equiv ' denotes distributional equivalence. Since the asymptotic limit of $\hat{\gamma}_{xx}^{k*}(h)$ is identical for all k , $\hat{\gamma}_{xx}^{k*}(h) = \hat{\gamma}_{xx}^*(h) + o_{a.s.}(1)$, and

$$\begin{aligned}
& \frac{1}{T^2} \sum_{t=1}^T \sum_{j=1}^q \sum_{k=1}^q x_{t-q+j} x'_{t-q+k} \\
\equiv & [q \hat{\gamma}_{xx}^*(0) + q o_{a.s.}(1)] + [(q-1) \hat{\gamma}_{xx}^*(1) + (q-1) o_{a.s.}(1)] + \dots + [\hat{\gamma}_{xx}^*(-(q-1))] \\
& + [\hat{\gamma}_{xx}^*(q-1)] + \dots + [(q-1) \hat{\gamma}_{xx}^*(1) + (q-1) o_{a.s.}(1)] \\
= & \sum_{h=-q+1}^{q-1} (q - |h|) \hat{\gamma}_{xx}^*(h) + q^2 o_{a.s.}(1).
\end{aligned}$$

Since $\hat{\gamma}_{xx}^*(h) \xrightarrow{a.s.} \int_0^1 J_C^* J_C^{*'}$ for all h , and $\frac{1}{q} \sum_{h=-q+1}^{q-1} \left(1 - \frac{|h|}{q}\right) = 1$, it follows that as $q, T \rightarrow \infty$,

$$\frac{1}{q^2 T^2} \sum_{t=1}^T \sum_{j=1}^q \sum_{k=1}^q x_{t-q+j} x'_{t-q+k} \equiv \frac{1}{q} \sum_{h=-q+1}^{q-1} \left(1 - \frac{|h|}{q}\right) \hat{\gamma}_{xx}^*(h) + o_{a.s.}(1) \xrightarrow{a.s.} \int_0^1 J_C^* J_C^{*'},$$

by Toeplitz's lemma and the assumption that $q/T = o(1)$. On the original probability space, therefore,

$$\frac{1}{q^2 T^2} \sum_{t=1}^T \sum_{j=1}^q \sum_{k=1}^q x_{t-q+j} x'_{t-q+k} \Rightarrow \int_0^1 J_C J'_C,$$

as $q, T \rightarrow \infty$. Similarly, let $\gamma_{ux}(h) = \int_0^1 dB_1 J'_C$, $\hat{\gamma}_{ux}^k(h) = \frac{1}{T} \sum_{t=q}^{T-q+1} u_{t-q+k} x'_{t-q+k-h}$, and $\hat{\gamma}_{ux}^q(h) = \hat{\gamma}_{ux}(h) = \frac{1}{T} \sum_{t=q}^{T-q+1} u_t x'_{t-h}$. For any fixed k and h , $\hat{\gamma}_{ux}^k(h) \Rightarrow \gamma_{ux}(h) = \int_0^1 dB_1 J'_C$. Using the same methods as for the denominator, as $q, T \rightarrow \infty$,

$$\frac{1}{q^2 T} \sum_{t=1}^T \sum_{j=1}^q \sum_{k=1}^q u_{t+j} x'_{t-q+k} \Rightarrow \int_0^1 dB_1 J'_C.$$

Combining these results,

$$T \left(\hat{\beta}^B(q) - 0 \right) \Rightarrow \int_0^1 dB_1 J'_C \left(\int_0^1 J_C J'_C \right)^{-1}.$$

2. (a) Next, consider the properties of $\hat{\beta}^U(q)$ under the alternative hypothesis. By summing up on both sides in equation (1),

$$\begin{aligned} r_{t+q}(q) &= \beta(x_t + x_{t+1} + \dots + x_{t+q-1}) + u_{t+q}(q) \\ &= \beta \left((x_t + Ax_t + \dots + A^{q-1}x_t) + \left(v_{t+1} + (Av_{t+1} + v_{t+2}) + \dots + \sum_{p=2}^q A^{q-p}v_{t+p-1} \right) \right) + u_{t+q}(q) \\ &= \beta^U(q) x_t + \beta \sum_{j=1}^{q-1} \sum_{p=q-j+1}^q A^{q-p} v_{t+p-q+j} + u_{t+q}(q), \end{aligned}$$

where $\beta^U(q) = \beta(I + A + \dots + A^{q-1})$. Thus,

$$\begin{aligned} &\frac{T}{q^2} \left(\hat{\beta}^U(q) - \beta^U(q) \right) \\ &= \left(\beta \frac{1}{q^2 T} \sum_{t=1}^T \sum_{j=1}^{q-1} \sum_{p=q-j+1}^q A^{q-p} v_{t+p-q+j} x'_t + \frac{1}{q^2 T} \sum_{t=1}^T u_{t+q}(q) x'_t \right) \left(\frac{1}{T^2} \sum_{t=1}^T x_t x'_t \right)^{-1}. \end{aligned}$$

By previous results, as $q, T \rightarrow \infty$, $\frac{1}{qT} \sum_{t=1}^T u_{t+q}(q) x'_t \Rightarrow \int_0^1 dB_1 J'_C$, so that $\frac{1}{q^2 T} \sum_{t=1}^T u_{t+q}(q) x'_t =$

$O_p(q^{-1})$. Further,

$$\begin{aligned}
& \frac{1}{q^2 T} \sum_{t=1}^T \sum_{j=1}^{q-1} \sum_{p=q-j+1}^q A^{q-p} v_{t+p-q+j} x'_t \\
&= \frac{1}{q^2 T} \sum_{t=1}^T \left(\left(\sum_{p=2}^q A^{q-p} \right) v_{t+1} + \dots + \left(\sum_{p=q}^q A^{q-p} \right) v_{t+q-1} \right) x'_t \\
&= \frac{1}{q} \left[\left(\frac{1}{q} \sum_{p=2}^q A^{q-p} \right) \hat{\gamma}_{vx}(1) + \dots + \left(\frac{1}{q} \sum_{p=q}^q A^{q-p} \right) \hat{\gamma}_{vx}(q-1) \right],
\end{aligned}$$

where $\hat{\gamma}_{vx}(h) = \frac{1}{T} \sum_{t=1}^T v_t x'_{t-h}$. By the local-to-unity property of A ,

$$\frac{1}{q} \sum_{p=h}^q A^{q-p} = \frac{1}{q} \sum_{p=h}^q \left(I + \frac{C}{T} \right)^{q-p} = \frac{1}{q} \sum_{p=h}^q \left(I + O\left(\frac{q}{T}\right) \right) = \frac{q-h+1}{q} I + O(qT^{-1}).$$

It follows that

$$\begin{aligned}
& \frac{1}{q^2 T} \sum_{t=1}^T \sum_{j=1}^{q-1} \sum_{p=q-j+1}^q A^{q-p} v_{t+p-q+j} x'_t \\
&= \frac{1}{q} \left[\left(\frac{q-1}{q} + O(qT^{-1}) \right) \hat{\gamma}_{vx}(1) + \dots + \left(\frac{1}{q} + O(qT^{-1}) \right) \hat{\gamma}_{vx}(q-1) \right] \\
&= \frac{1}{q} \sum_{h=1}^{q-1} \left(1 - \frac{h}{q} \right) \hat{\gamma}_{vx}(h) + O_p(qT^{-1}).
\end{aligned}$$

For any h ,

$$\hat{\gamma}_{vx}(h) = \frac{1}{T} \sum_{t=1}^T v_t x'_{t-h} \Rightarrow \int_0^1 dB_2 J'_C + \Lambda_{22}(h) = \gamma_{vx}(h) + \Lambda_{22}(h),$$

where $\Lambda_{22}(h) = \sum_{k=h}^{\infty} E(v_{i,k} v'_{i,0})$. Since $\frac{1}{q} \sum_{h=1}^{q-1} \left(1 - \frac{h}{q} \right) = \frac{1}{2}$, as $q \rightarrow \infty$, and $\Lambda_{22}(h) \rightarrow 0$ as $h \rightarrow \infty$,

it follows that

$$\frac{2}{q} \sum_{h=1}^{q-1} \left(1 - \frac{h}{q} \right) \Lambda_{22}(h) \rightarrow \Lambda_{22}(1) = \Lambda_{22},$$

as $q \rightarrow \infty$. Thus as $q, T \rightarrow \infty$,

$$\frac{2}{q} \sum_{h=1}^{q-1} \left(1 - \frac{h}{q} \right) \hat{\gamma}_{vx}(h) \Rightarrow \int_0^1 dB_2 J'_C + \Lambda_{22}.$$

Summing up,

$$\frac{2T}{q^2} \left(\hat{\beta}^U(q) - \beta^U(q) \right) \Rightarrow \beta \left(\int_0^1 dB_2 J'_C + \Lambda_{22} \right) \left(\int_0^1 J_C J'_C \right).$$

(b) As in part 2.(a),

$$r_{t+q}(q) = \beta(x_t + x_{t+1} + \dots + x_{t+q-1}) + u_{t+q}(q) = \beta x_{t+q-1}(q) + u_{t+q}(q).$$

Thus,

$$\hat{\beta}^B(q) = \left(\beta \sum_{t=1}^T x_{t+q-1}(q) x_t(q)' + \sum_{t=1}^T u_{t+q}(q) x_t(q)' \right) \left(\sum_{t=1}^T x_t(q) x_t(q)' \right)^{-1}.$$

Observe that $x_{t+q-1} = A^{q-1}x_t + \sum_{p=2}^q A^{q-p}v_{t+p-1}, \dots, x_t = A^{q-1}x_{t-q+1} + \sum_{p=2}^q A^{q-p}v_{t+p-q}$, and one can therefore write

$$\begin{aligned} x_{t+q-1}(q) &= x_{t+q-1} + \dots + x_{t+1} + x_t \\ &= A^{q-1}(x_t + \dots + x_{t-q+2} + x_{t-q+1}) + \sum_{p=2}^q A^{q-p}(v_{t+p-1} + \dots + v_{t+p-q+1} + v_{t+p-q}) \\ &= A^{q-1}x_t(q) + \sum_{p=2}^q A^{q-p}v_{t+p-1}(q). \end{aligned}$$

Thus,

$$\sum_{t=1}^T x_{t+q-1}(q) x_t(q)' = A^{q-1} \sum_{t=1}^T x_t(q) x_t(q)' + \sum_{t=1}^T \sum_{p=2}^q A^{q-p} v_{t+p-1}(q) x_t(q)',$$

and

$$\hat{\beta}^B(q) = \beta A^{q-1} + \left(\beta \sum_{t=1}^T \sum_{p=2}^q A^{q-p} v_{t+p-1}(q) x_t(q)' + \sum_{t=1}^T u_{t+q}(q) x_t(q)' \right) \left(\sum_{t=1}^T x_t(q) x_t(q)' \right)^{-1}.$$

Let $\beta^B(q) = \beta A^{q-1}$ and consider

$$\begin{aligned} &\frac{T}{q} \left(\hat{\beta}^B(q) - \beta^B(q) \right) \\ &= \left(\beta \frac{1}{q^3 T} \sum_{t=1}^T \sum_{p=2}^q A^{q-p} v_{t+p-1}(q) x_t(q)' + \frac{1}{q^3 T} \sum_{t=1}^T u_{t+q}(q) x_t(q)' \right) \left(\frac{1}{q^2 T^2} \sum_{t=1}^T x_t(q) x_t(q)' \right)^{-1}. \end{aligned}$$

Since, as $q, T \rightarrow \infty$, $\frac{1}{q^2 T} \sum_{t=1}^T u_{t+q}(q) x_t(q)' \Rightarrow \int_0^1 dB_1 J_C'$, it follows that $\frac{1}{q^3 T} \sum_{t=1}^T u_{t+q}(q) x_t(q)' = o_p(q^{-1})$. Next,

$$\frac{1}{q^3 T} \sum_{t=1}^T \sum_{p=2}^q A^{q-p} v_{t+p-1}(q) x_t(q)' = \frac{1}{q} \sum_{p=2}^q A^{q-p} \frac{1}{q^2 T} \sum_{t=1}^T v_{t+p-1}(q) x_t(q)'$$

Again, let $\hat{\gamma}_{vx}(h) = T^{-1} \sum_{t=1}^T v_t x'_{t-h}$ and use the Skorohod construction, such that

$$\hat{\gamma}_{vx}^*(h) \xrightarrow{a.s.} \gamma_{vx}^*(h) + \Lambda_{22}(h) = \int_0^1 dB_2^* J_C^{*'} + \Lambda_{22}(h)$$

where $\hat{\gamma}_{vx}^*(h) \equiv \hat{\gamma}_{vx}(h)$ and $\int_0^1 dB_2^* J_C^{*'} \equiv \int_0^1 dB_2 J_C'$. By similar arguments as in 1.(b),

$$\frac{1}{q} \sum_{p=2}^q A^{q-p} \frac{1}{q^2 T} \sum_{t=1}^T v_{t+p-1}(q) x_t(q)' \equiv \frac{1}{q} \sum_{p=2}^q A^{q-p} \frac{1}{q} \sum_{h=-q+1}^{q-1} \left(1 - \frac{|h|}{q}\right) \hat{\gamma}_{vx}^*(p-1+h) + o_{a.s.}(1).$$

Since $\Lambda_{22}(h) \rightarrow 0$ as $h \rightarrow \infty$ and $\Lambda_{22}(h) \rightarrow \Omega_{22} = \sum_{k=-\infty}^{\infty} E(v_{i,k} v'_{i,0})$ as $h \rightarrow -\infty$, it follows that

$$\frac{1}{q} \sum_{h=-q+1}^{q-1} \left(1 - \frac{|h|}{q}\right) \Lambda_{22}(p-1+h) \rightarrow \Omega_{22}$$

as $q \rightarrow \infty$, for any p , and thus

$$\frac{1}{q} \sum_{p=2}^q A^{q-p} \frac{1}{q} \sum_{h=-q+1}^{q-1} \left(1 - \frac{|h|}{q}\right) \hat{\gamma}_{vx}^*(p-1+h) + o_{a.s.}(1) \xrightarrow{a.s.} \int_0^1 dB_2^* J_C^{*'} + \Omega_{22}.$$

On the original probability space

$$\frac{1}{q} \sum_{p=2}^q A^{q-p} \frac{1}{q^2 T} \sum_{t=1}^T v_{t+p-1}(q) x_t(q)' \Rightarrow \int_0^1 dB_2 J_C' + \Omega_{22},$$

and

$$\frac{T}{q} \left(\hat{\beta}^B(q) - \beta^B(q) \right) \Rightarrow \beta \left(\int_0^1 dB_2 J_C' + \Omega_{22} \right) \left(\int_0^1 J_C J_C' \right)^{-1}.$$

■

Proof of Theorem 2. Only the derivations for $\hat{\beta}^{B^+}(q)$ are shown. The proof for $\hat{\beta}^{U^+}(q)$ follows in a similar manner, although with somewhat simpler arguments given the short-run nature of the regressors in this case.

1. Start with $\hat{\beta}^{B^+}(q)$ under the null-hypothesis. Let $\mathbf{r}_{+q}^q = (r_{1+q}(q), \dots, r_T(q))'$ be the $T \times 1$ vector of observations, and define \mathbf{x}^q and \mathbf{v}_{+q}^q analogously. The OLS estimator of $\beta^B(q)$ in (15) is now given by

$$\hat{\beta}^{B^+}(q) = (\mathbf{r}_{+q}^{q'} Q_{\mathbf{v}^q} \mathbf{x}^q) (\mathbf{x}^{q'} Q_{\mathbf{v}^q} \mathbf{x}^q)^{-1}.$$

where $Q_{\mathbf{v}^q} = I - \mathbf{v}_{+q}^q (\mathbf{v}_{+q}^{q'} \mathbf{v}_{+q}^q)^{-1} \mathbf{v}_{+q}^{q'}$. Under the null-hypothesis,

$$Q_{\mathbf{v}^q} \mathbf{r}_{+q}^q = Q_{\mathbf{v}^q} (\mathbf{v}_{+q}^q \gamma' + (\mathbf{u}_{+q}^q - \mathbf{v}_{+q}^q \gamma')) = Q_{\mathbf{v}^q} \mathbf{u}_{+q}^q$$

and

$$T \left(\hat{\beta}^{B+}(q) - 0 \right) = (q^{-2} T^{-1} \mathbf{u}_{+q}^{q'} Q_{\mathbf{v}^q} \mathbf{x}^q) (q^{-2} T^{-2} \mathbf{x}^{q'} Q_{\mathbf{v}^q} \mathbf{x}^q)^{-1}.$$

As $q, T \rightarrow \infty$,

$$\begin{aligned} (qT)^{-2} \mathbf{x}^{q'} Q_{\mathbf{v}^q} \mathbf{x}^q &= (qT)^{-2} \mathbf{x}^{q'} \mathbf{x}^q - qT^{-1} (q^{-2} T^{-1} \mathbf{x}^{q'} \mathbf{v}_{+q}^q) (q^{-1} T^{-1} \mathbf{v}_{+q}^{q'} \mathbf{v}_{+q}^q)^{-1} (q^{-2} T^{-1} \mathbf{v}_{+q}^{q'} \mathbf{x}^q) \\ &\Rightarrow \int_0^1 J_C J_C', \end{aligned}$$

since $q/T \rightarrow 0$. Next,

$$Q_{\mathbf{v}^q} \mathbf{u}_{+q}^q = \mathbf{u}_{+q}^q - \mathbf{v}_{+q}^q \left(\frac{1}{qT} \sum_{t=1}^T v_{t+q}(q) v_{t+q}(q)' \right)^{-1} \left(\frac{1}{qT} \sum_{t=1}^T u_{t+q}(q) v_{t+q}(q)' \right).$$

Let

$$\hat{\gamma}_{vv}(h) = \frac{1}{T} \sum_{t=q}^{T-q+1} v_t v_{t-h}', \text{ and } \hat{\gamma}_{uv}(h) = \frac{1}{T} \sum_{t=q}^{T-q+1} u_t v_{t-h}'$$

and define $\hat{\gamma}_{vv}^k(h)$ and $\hat{\gamma}_{uv}^k(h)$ in an analogous manner to the previous proof. Similarly to above,

$$\frac{1}{qT} \sum_{t=1}^T v_{t+q}(q) v_{t+q}(q)' = \sum_{h=-q+1}^{q-1} \left(1 - \frac{|h|}{q} \right) \hat{\gamma}_{vv}(h) + O_p(qT^{-1/2}) \rightarrow_p \Omega_{22},$$

as $q, T \rightarrow \infty$, by standard results (e.g. Andrews, 1991), since $qT^{-1/2} = o(1)$. By identical arguments,

as $q, T \rightarrow \infty$,

$$\frac{1}{qT} \sum_{t=1}^T u_{t+q}(q) v_{t+q}(q)' = \sum_{h=-q+1}^{q-1} \left(1 - \frac{|h|}{q} \right) \hat{\gamma}_{uv}(h) + O_p(qT^{-1/2}) \rightarrow_p \omega_{12}.$$

Again, using the same methods as in the proof of part 1.(b) in Theorem 1, it follows that

$$\frac{1}{q^2 T} \mathbf{v}_{+q}^{q'} \mathbf{x}^q = \frac{1}{q^2 T} \sum_{t=1}^T v_{t+q}(q) x_t(q)' \Rightarrow \int_0^1 dB_2 J_C'.$$

Thus,

$$\begin{aligned} q^{-2}T^{-1}\mathbf{u}_{+q}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q &= q^{-2}T^{-1}\mathbf{u}_{+q}^{q'}\mathbf{x}^q - (q^{-1}T^{-1}\mathbf{u}_{+q}^{q'}\mathbf{v}_{+q}^q)(q^{-1}T^{-1}\mathbf{v}_{+q}^{q'}\mathbf{v}_{+q}^q)^{-1}(q^{-2}T^{-1}\mathbf{v}_{+q}^{q'}\mathbf{x}^q) \\ \Rightarrow \int_0^1 dB_1 J'_C - \omega_{12}\Omega_{22}^{-1} \int_0^1 dB_2 J'_C &= \int_0^1 dB_{1.2} J'_C, \end{aligned}$$

and

$$T\left(\hat{\beta}^{B+}(q) - 0\right) = (q^{-2}T^{-1}\mathbf{u}_{+q}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)(q^{-2}T^{-2}\mathbf{x}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)^{-1} \Rightarrow \int_0^1 dB_{1.2} J'_C \left(\int_0^1 J_C J'_C\right)^{-1}.$$

2. Now consider $\hat{\beta}^{B+}(q)$ under the alternative hypothesis. Observe that,

$$Q_{\mathbf{v}^q}\mathbf{r}_{+q}^q = Q_{\mathbf{v}^q}(\mathbf{x}_{+q-1}^q\beta' + \mathbf{v}_{+q}^q\gamma' + (\mathbf{u}_{+q}^q - \mathbf{v}_{+q}^q\gamma')) = Q_{\mathbf{v}^q}\mathbf{x}_{+q-1}^q\beta' + Q_{\mathbf{v}^q}\mathbf{u}_{+q}^q,$$

and

$$\hat{\beta}^{B+}(q) = \beta(\mathbf{x}_{+q-1}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)(\mathbf{x}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)^{-1} + (\mathbf{u}_{+q}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)(\mathbf{x}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)^{-1}.$$

Recall that $x_{t+q-1} = A^{q-1}x_t(q) + \sum_{p=2}^q A^{q-p}v_{t+p-1}(q)$. Thus

$$\mathbf{x}_{+q-1}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q = \left(A^{q-1}\mathbf{x}^{q'} + \sum_{p=2}^q A^{q-p}\mathbf{v}_{+p-1}^{q'}\right)Q_{\mathbf{v}^q}\mathbf{x}^q = A^{q-1}\mathbf{x}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q + \sum_{p=2}^q A^{q-p}\mathbf{v}_{+p-1}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q,$$

and

$$\hat{\beta}^{B+}(q) = \beta A^{q-1} + \beta \sum_{p=2}^q A^{q-p}(\mathbf{v}_{+p-1}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)(\mathbf{x}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)^{-1} + (\mathbf{u}_{+q}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)(\mathbf{x}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)^{-1}.$$

Again, let $\beta^B(q) = \beta A^{q-1}$ and write

$$\begin{aligned} &\frac{T}{q}\left(\hat{\beta}^{B+}(q) - \beta^B(q)\right) \\ &= \beta \frac{1}{q} \sum_{p=2}^q A^{q-p} (q^{-2}T^{-1}\mathbf{v}_{+p-1}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)(q^{-2}T^{-2}\mathbf{x}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)^{-1} \\ &\quad + q^{-1}(q^{-2}T^{-1}\mathbf{u}_{+q}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)(q^{-2}T^{-2}\mathbf{x}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)^{-1} \\ &= \beta \frac{1}{q} \sum_{p=2}^q A^{q-p} (q^{-2}T^{-1}\mathbf{v}_{+p-1}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)(q^{-2}T^{-2}\mathbf{x}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q)^{-1} + o_p(q^{-1}). \end{aligned}$$

Consider

$$q^{-2}T^{-1}\mathbf{v}_{+p-1}^{q'}Q_{\mathbf{v}^q}\mathbf{x}^q = q^{-1}T^{-1}\left[\left(q^{-1}\mathbf{v}_{+p-1}^{q'} - (q^{-1}T^{-1}\mathbf{v}_{+p-1}^{q'}\mathbf{v}_{+q}^q)(q^{-1}T^{-1}\mathbf{v}_{+q}^{q'}\mathbf{v}_{+q}^q)^{-1}q^{-1}\mathbf{v}_{+q}^{q'}\right)\right]\mathbf{x}^q.$$

By similar arguments as above, as $q, T \rightarrow \infty$, $q^{-1}T^{-1}\mathbf{v}_{+p-1}^{q'}\mathbf{v}_{+q}^q \rightarrow_p \Omega_{22}$ and $q^{-1}T^{-1}\mathbf{v}_{+q}^{q'}\mathbf{v}_{+q}^q \rightarrow_p \Omega_{22}$.

Further,

$$v_{t+p-1}(q) - v_{t+q}(q) = \sum_{h=0}^{q-p} v_{t-h} - \sum_{h=0}^{q-p} v_{t+q-h},$$

and

$$\begin{aligned} & \frac{1}{q^3T} \sum_{t=1}^T \sum_{p=2}^q A^{q-p} \left(\sum_{h=0}^{q-p} v_{t-h} - \sum_{h=0}^{q-p} v_{t+q-h} \right) x_t(q)' + o_p(1) \\ &= \frac{1}{q^3T} \sum_{t=1}^T \left(\sum_{p=2}^q \sum_{h=0}^{q-p} v_{t-h} - \sum_{p=2}^q \sum_{h=0}^{q-p} v_{t+q-h} \right) x_t(q)' + O_p(T^{-1}) \\ &= \frac{1}{q^2T} \sum_{t=1}^T \left(\sum_{h=0}^{q-2} \left(1 - \frac{h}{q}\right) v_{t-h} x_t(q)' - \sum_{h=0}^{q-2} \left(1 - \frac{h}{q}\right) v_{t+q-h} x_t(q)' \right) + O_p(T^{-1}) \\ &\Rightarrow \frac{1}{2} \int_0^1 dB_2 J'_C + \frac{1}{2} \int_0^1 dB_2 J'_C = \int_0^1 dB_2 J'_C, \end{aligned}$$

as $q, T \rightarrow \infty$. Thus,

$$\frac{T}{q} \left(\hat{\beta}^{B^+}(q) - \beta^B(q) \right) \Rightarrow \left(\int_0^1 dB_2 J_C \right) \left(\int_0^1 J_C J'_C \right)^{-1}.$$

■

Proof of Corollary 1. Under the null hypothesis, following the arguments in the Proof of Theorem 2,

$$\begin{aligned} \omega_{11.2}^U &= \frac{1}{qT} \sum_{t=1}^T \hat{u}_t^{U+}(q)^2 = \frac{1}{qT} \sum_{t=1}^T \left(u_{t+q.2}(q) + O_p\left(\frac{q}{T}\right) \right)^2 \\ &= \frac{1}{qT} \sum_{t=1}^T \left(u_{t+q}(q) - \omega_{12}\Omega_{22}^{-1}v_{t+q}(q) \right)^2 + O_p\left(\frac{q}{T}\right) \xrightarrow{p} \omega_{11.2}. \end{aligned}$$

Identical arguments hold for $\hat{\omega}_{11.2}^B$. The following t -statistics can now be formed

$$\frac{\frac{T}{q}\hat{\beta}^{U+}(q)}{\sqrt{\hat{\omega}_{11.2}^U T^{-2} a' \left(\sum_{t=1}^T \underline{z}_t \underline{z}_t'\right)^{-1} a}} = \frac{\hat{\beta}^{U+}(q)}{\sqrt{q^2 \hat{\omega}_{11.2}^U a' \left(\sum_{t=1}^T \underline{z}_t \underline{z}_t'\right)^{-1} a}} = t^{U+}(q) / \sqrt{q},$$

and

$$\frac{T\hat{\beta}^{B+}(q)}{\sqrt{\hat{\omega}_{11.2}^B (qT)^{-1} a' \left(\sum_{t=1}^T \underline{z}_t(q) \underline{z}_t(q)'\right)^{-1} a}} = \frac{\hat{\beta}^{B+}(q)}{\sqrt{q^2 \hat{\omega}_{11.2}^B a' \left(\sum_{t=1}^T \underline{z}_t(q) \underline{z}_t(q)'\right)^{-1} a}} = t^{B+}(q) / \sqrt{q},$$

where $t^{U+}(q)$ and $t^{B+}(q)$ are the standard t -statistics corresponding to $\hat{\beta}^{U+}(q)$ and $\hat{\beta}^{B+}(q)$. The results now follow directly from the asymptotically mixed normal distributions of the estimators. ■

References

- Andrews, D.W.K., 1991. Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica* 59, 817-858.
- Ang, A., and G. Bekaert, 2003. Stock Return Predictability: Is it There? Working paper, Columbia University.
- Boudoukh J., and M. Richardson, 1993. Stock Returns and Inflation: A Long-Horizon Perspective, *American Economic Review* 83, 1346-1355.
- Boudoukh J., M. Richardson, and R.F. Whitelaw, 2005. The myth of long-horizon predictability, NBER working paper 11841.
- Berkowitz, J., and L. Giorgianni, 2001. Long-horizon exchange rate predictability?, *Review of Economics and Statistics* 83, 81-91.
- Campbell, J.Y., 2001. Why long horizons? A study of power against persistent alternatives, *Journal of Empirical Finance* 8, 459-491.
- Campbell, J.Y., and R. Shiller, 1988. Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661-676.
- Campbell, J.Y., and M. Yogo, 2005. Efficient Tests of Stock Return Predictability, forthcoming *Journal of Financial Economics*.
- Cavanagh, C., G. Elliot, and J. Stock, 1995. Inference in models with nearly integrated regressors, *Econometric Theory* 11, 1131-1147.
- Corbae D., S. Ouliaris, and P.C.B. Phillips, 2002. Band Spectral Regression with Trending Data, *Econometrica* 70, 1067-1109.
- Daniel, K., 2001. The power and size of mean reversion tests, *Journal of Empirical Finance* 8, 493-535.
- Fama, E.F., and K.R. French, 1988a. Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3-25.

Fama, E.F., and K.R. French, 1988b. Permanent and Temporary Components of Stock Prices, *Journal of Political Economy*, 96, 256-73.

Fisher, M.E., and J.J. Seater, 1993. Long-Run Neutrality and Superneutrality in an ARIMA Framework, *American Economic Review* 83, 402-415.

Goetzman W.N., and P. Jorion, 1993. Testing the Predictive Power of Dividend Yields, *Journal of Finance* 48, 663-679.

Hansen, L.P., and R.J. Hodrick, 1980. Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis, *Journal of Political Economy* 88, 829-853.

Hjalmarsson, E., 2005. Fully modified estimation with nearly integrated regressors, Working paper, Division of International Finance, Federal Reserve Board.

Hodrick, R.J., 1992. Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, *Review of Financial Studies* 5, 357-386.

Jansson, M., and M.J. Moreira, 2004. Optimal Inference in Regression Models with Nearly Integrated Regressors, Working Paper, Harvard University.

Lanne, M., 2002. Testing the Predictability of Stock Returns, *Review of Economics and Statistics* 84, 407-415.

Lewellen, J., 2004. Predicting returns with financial ratios, *Journal of Financial Economics*, 74, 209-235.

Mankiw, N.G., and M.D. Shapiro, 1986. Do we reject too often? Small sample properties of tests of rational expectations models, *Economic Letters* 20, 139-145.

Mark, N.C., 1995. Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability, *American Economic Review* 85, 201-218.

Mark, N.C., and D. Sul, 2004. The Use of Predictive Regressions at Alternative Horizons in Finance and Economics, NBER Technical Working Paper 298.

Mishkin, F.S., 1990. What does the term structure tell us about future inflation?, *Journal of Monetary Economics* 25, 77-95.

Mishkin, F.S., 1992. Is the Fisher effect for real?, *Journal of Monetary Economics* 30, 195-215.

Moon, R., A. Rubia, and R. Valkanov, 2004. Long-Horizon Regressions when the Predictor is Slowly Varying, Working Paper, UCLA, Anderson School of Management.

Nelson, C.R., and M.J. Kim, 1993. Predictable Stock Returns: The Role of Small Sample Bias, *Journal of Finance* 48, 641-661.

Newey, W., and K. West, 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703-708.

Phillips, P.C.B, 1987. Towards a Unified Asymptotic Theory of Autoregression, *Biometrika* 74, 535-547.

Phillips, P.C.B, 1988. Regression Theory for Near-Integrated Time Series, *Econometrica* 56, 1021-1043.

Phillips, P.C.B, 1991a. Optimal Inference in Cointegrated Systems, *Econometrica* 59, 283-306.

Phillips, P.C.B, 1991b. Spectral regression for cointegrated time series, in *Nonparametric and Semiparametric Methods in Economics and Statistics*, edited by W. Barnett, J. Powell, and G. Tauchen. Cambridge, Cambridge University Press.

Phillips, P.C.B, 1995. Fully Modified Least Squares and Vector Autoregression, *Econometrica* 63, 1023-1078.

Phillips, P.C.B, and B. Hansen, 1990. Statistical Inference in Instrumental Variables Regression with I(1) Processes, *Review of Economic Studies* 57, 99-125.

Phillips, P.C.B., and V. Solo, 1992. Asymptotics for Linear Processes, *Annals of Statistics*, 20, 971-1001.

Polk, C., S. Thompson, and T. Vuolteenaho, 2004. Cross-sectional Forecasts of the Equity Premium, Working Paper, Harvard University.

Rapach D.E., and M.E. Wohar, 2005. Valuation ratios and long-horizon stock price predictability, *Journal of Applied Econometrics* 20, 327-344.

Richardson, M., and T. Smith, 1991. Tests of Financial Models in the Presence of Overlapping Observations, *Review of Financial Studies* 4, 227-254.

Richardson, M., and J.H. Stock, 1989. Drawing Inferences from Statistics Based on Multiyear Asset Returns, *Journal of Financial Economics* 25, 323-348.

Rossi, B., 2005. Testing long-horizon predictive ability with high persistence, and the Meese-Rogoff puzzle, *International Economic Review* 46, 61-92.

Stambaugh, R., 1999. Predictive regressions, *Journal of Financial Economics* 54, 375-421.

Stock, J.H., 1991. Confidence intervals for the largest autoregressive root in U.S. economic time-series. *Journal of Monetary Economics* 28, 435-460.

Torous, W., R. Valkanov, and S. Yan, 2004. On Predicting Stock Returns with Nearly Integrated Explanatory Variables, *Journal of Business* 77, 937-966.

Valkanov, R., 2003. Long-horizon regressions: theoretical results and applications, *Journal of Financial Economics* 68, 201-232.

Table 1: Characteristics of the predictor variables. This table reports the key time-series characteristics of the dividend-price ratio ($d - p$) and the earnings-price ratio ($e - p$). The S&P 500 variables are on an annual frequency and the CRSP data on a monthly frequency. The column labeled δ gives the estimated correlations between the innovations to the predictor variables and the innovations to the corresponding excess returns. The last column gives a 95% confidence interval of the local-to-unity parameter C , obtained by inverting the DF-GLS unit-root test statistic.

Series	Sample Period	Obs.	Variable	δ	95% CI for C
S&P 500	1880 – 2002	123	$d - p$	-0.845	[-6.107, 4.020]
	1880 – 2002	123	$e - p$	-0.962	[-28.262, -4.232]
CRSP	1926 – 2002	913	$d - p$	-0.950	[-12.683, 2.377]
	1926 – 2002	913	$e - p$	-0.987	[-14.797, 1.711]

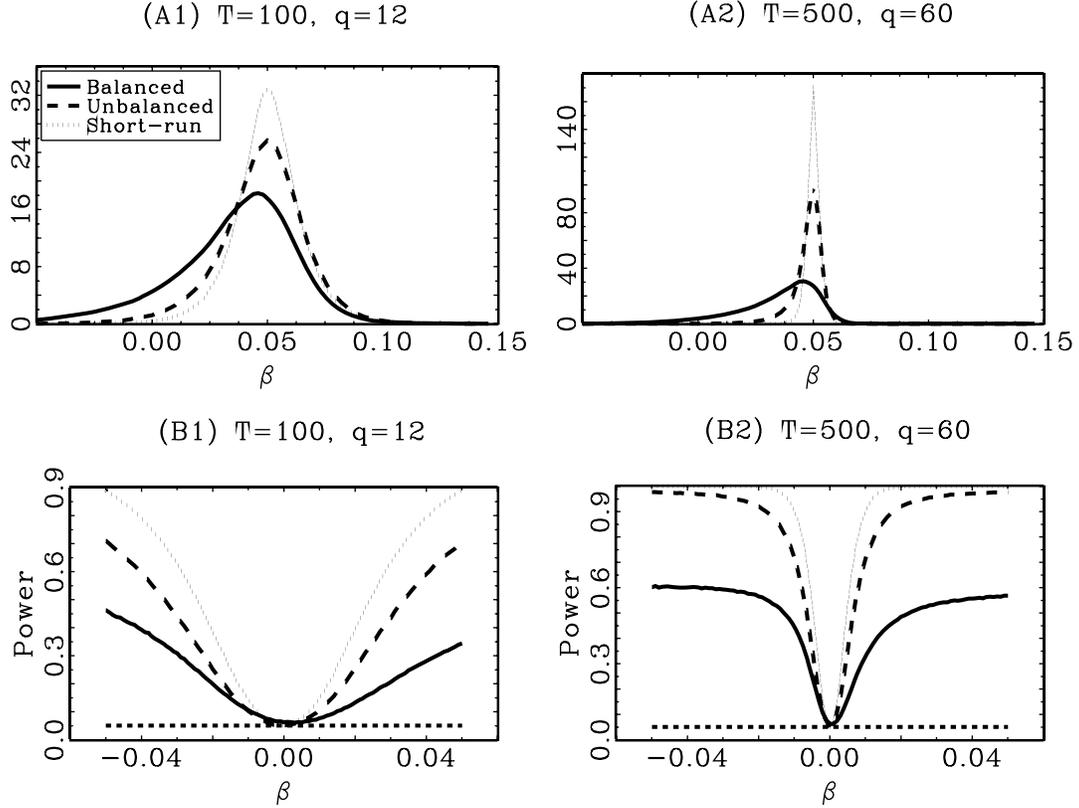


Figure 1: Results from the Monte Carlo simulation for $C = 0$. The top two graphs, (A1) and (A2), show the kernel estimates of the densities of the estimated coefficients, using the estimators $\hat{\beta}^{B^+}(q)$, $\hat{\beta}^{U^+}(q)$, and $\hat{\beta}^{U^+}(q=1)$, referred to as Balanced, Unbalanced and Short-run, respectively, in the legend. The true value of β is equal to 0.05. As suggested by the result in Theorem 1, $\hat{\beta}^{U^+}(q)$ and $\hat{\beta}^{B^+}(q)$ are divided by $(1 + A + \dots + A^{q-1})$ and A^{q-1} , respectively, with $A = 1 + C/T$, to enable a fair comparison between the estimators. The bottom graphs, (B1) and (B2), show the average rejection rates of the scaled t -tests corresponding to the respective estimators; the flat dashed lines show the 5% level.

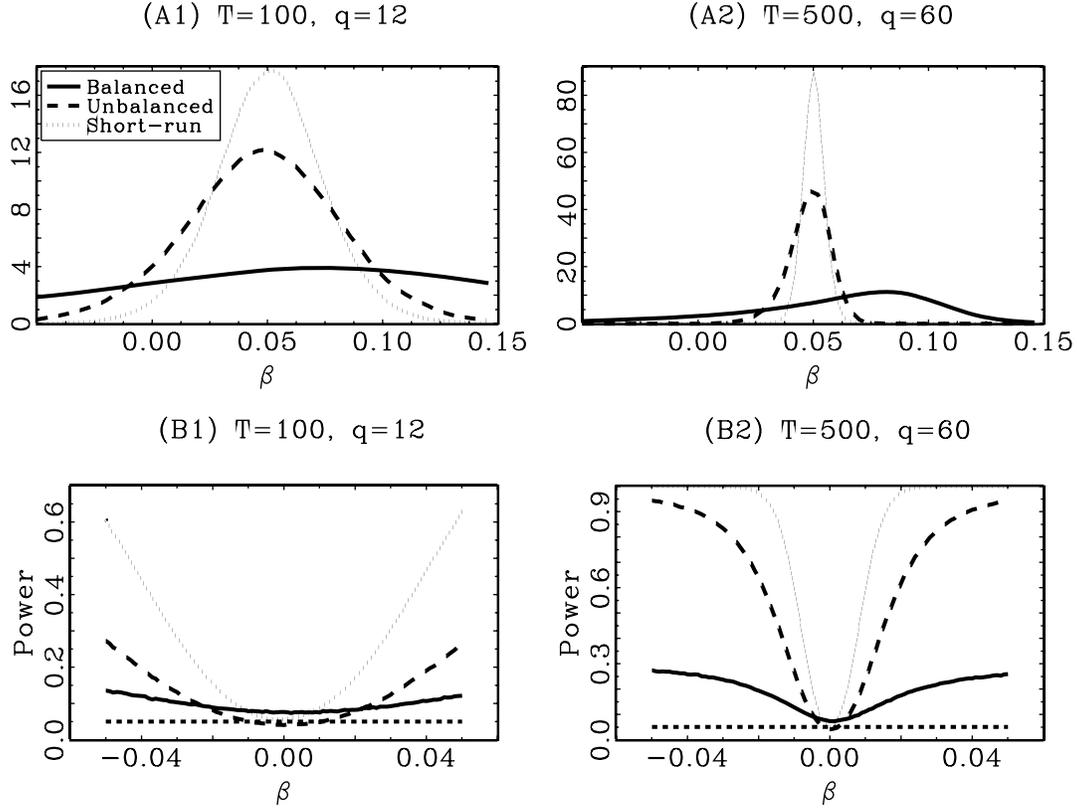


Figure 2: Results from the Monte Carlo simulation for $C = -10$. The top two graphs, (A1) and (A2), show the kernel estimates of the densities of the estimated coefficients, using the estimators $\hat{\beta}^{B^+}(q)$, $\hat{\beta}^{U^+}(q)$, and $\hat{\beta}^{U^+}(q=1)$, referred to as Balanced, Unbalanced and Short-run, respectively, in the legend. The true value of β is equal to 0.05. As suggested by the result in Theorem 1, $\hat{\beta}^{U^+}(q)$ and $\hat{\beta}^{B^+}(q)$ are divided by $(1 + A + \dots + A^{q-1})$ and A^{q-1} , respectively, with $A = 1 + C/T$, to enable a fair comparison between the estimators. The bottom graphs, (B1) and (B2), show the average rejection rates of the scaled t -tests corresponding to the respective estimators; the flat dashed lines show the 5% level.

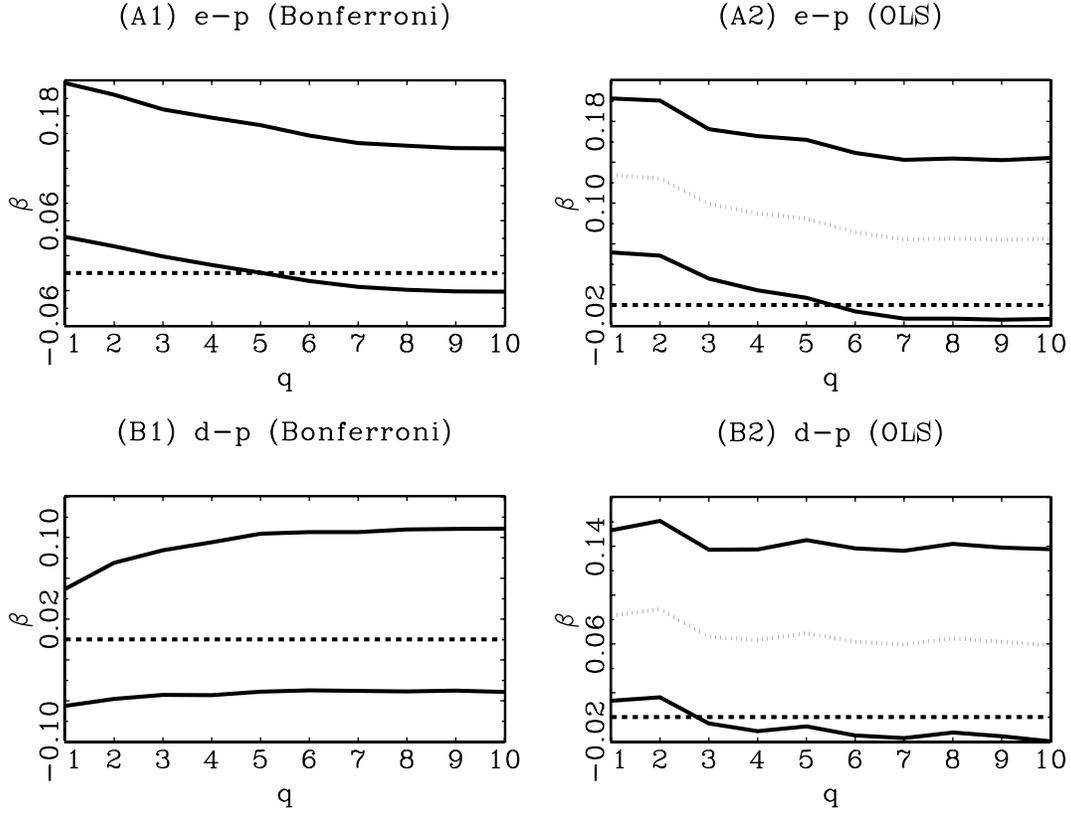


Figure 3: Estimation results for the annual S&P 500 data. The top two graphs show the results for the earnings-price ratio ($e - p$). In the left hand one, (A1), the 90% Bonferroni confidence intervals for the slope coefficient β are plotted as a function of the forecasting horizon q . These are obtained from the endogeneity corrected long-run estimator $\hat{\beta}^{U+}(q)$ and the corresponding scaled t -statistics. They are thus asymptotically valid for statistical inference. To enable comparison across q , the confidence bounds have been scaled by q^{-1} as suggested by Theorem 1. As a comparison, in the right hand graph, (A2), the standard 90% confidence intervals for the long-run OLS estimator $\hat{\beta}^U(q)$ are shown, along with the point estimate $\hat{\beta}^U(q)$ (the dotted line). These confidence intervals are calculated as $\hat{\beta}^U(q) \pm 1.65 \times \hat{\beta}^U(q) \sqrt{q}/t^U(q)$ and are not valid since they are based on the assumption of an asymptotically mixed normal distribution of $\hat{\beta}^U(q)$. Again, the point estimates and the confidence bounds have been scaled by q^{-1} . The flat dashed lines indicate zero. The bottom graphs, (B1) and (B2), show the corresponding results for the dividend-price ratio ($d - p$).

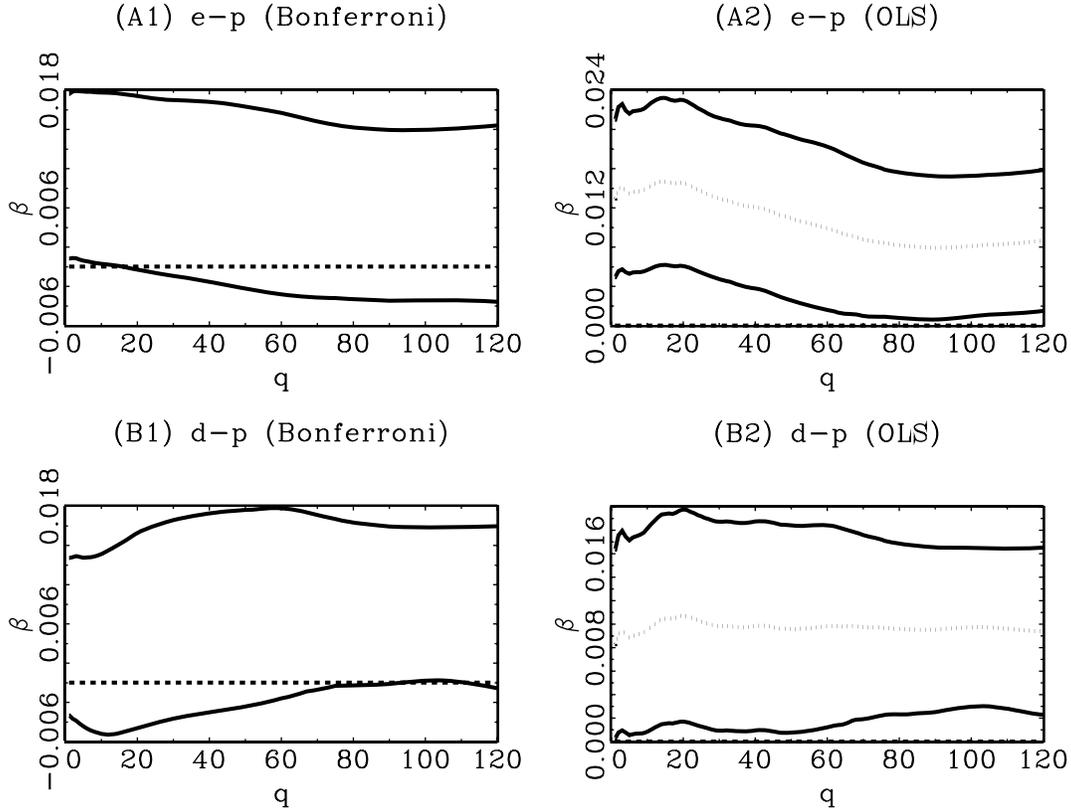


Figure 4: Estimation results for the monthly CRSP data. The top two graphs show the results for the earnings-price ratio ($e - p$). In the left hand one, (A1), the 90% Bonferroni confidence intervals for the slope coefficient β are plotted as a function of the forecasting horizon q . These are obtained from the endogeneity corrected long-run estimator $\hat{\beta}^{U+}(q)$ and the corresponding scaled t -statistics. They are thus asymptotically valid for statistical inference. To enable comparison across q , the confidence bounds have been scaled by q^{-1} as suggested by Theorem 1. As a comparison, in the right hand graph, (A2), the standard 90% confidence intervals for the long-run OLS estimator $\hat{\beta}^U(q)$ are shown, along with the point estimate $\hat{\beta}^U(q)$ (the dotted line). These confidence intervals are calculated as $\hat{\beta}^U(q) \pm 1.65 \times \hat{\beta}^U(q) \sqrt{q}/t^U(q)$ and are thus not valid since they are based on the assumption of an asymptotically mixed normal distribution of $\hat{\beta}^U(q)$. Again, the point estimates and the confidence bounds have been scaled by q^{-1} . The flat dashed lines indicate zero. The bottom graphs, (B1) and (B2), show the corresponding results for the dividend-price ratio ($d - p$).

International Finance Discussion Papers

<u>IFDP Number</u>	<u>Titles</u>	<u>Author(s)</u>
2006		
851	Exchange-Rate Pass-Through in the G-7 Countries	Jane E. Ihrig Mario Marazzi Alexander D. Rothenberg
850	The Adjustment of Global External Imbalances: Does Partial Exchange Rate Pass-Through to Trade Prices Matter?	Christopher Gust Nathan Sheets
2005		
849	Interest Rate Rules, Endogenous Cycles and Chaotic Dynamics in Open Economies	Marco Airaud Luis-Felipe Zanna
848	Fighting Against Currency Depreciation Macroeconomic Instability and Sudden Stops	Luis-Felipe Zanna
847	The Baby Boom Predictability in House Prices and Interest Rates	Robert F. Martin
846	Explaining the Global Pattern of Current Account Imbalances	Joseph W. Gruber Steven B. Kamin
845	DSGE Models of High Exchange-Rate Volatility and Low Pass-Through	Giancarlo Corsetti Luca Dedola Sylvain Leduc
844	The Response of Global Equity Indexes to U.S. Monetary Policy Announcements	Jon Wongswan
843	Accounting Standards and Information: Inferences from Cross-Listed Financial Firms	John Ammer Nathanael Clinton Gregory P. Nini
842	Alternative Procedures for Estimating Vector Autoregressions Identified with Long-Run Restrictions	Lawrence J Christiano Martin Eichenbaum Robert J. Vigfusson

Please address requests for copies to International Finance Discussion Papers, Publications, Stop 127, Board of Governors of the Federal Reserve System, Washington, DC 20551.
Email: publications-bog@frb.gov. Fax (202) 728-5886.

International Finance Discussion Papers

<u>IFDP Number</u>	<u>Titles</u>	<u>Author(s)</u>
841	Monetary Policy and House Prices: A Cross-Country Study	Alan G. Ahearne John Ammer Brian M. Doyle Linda S. Kole Robert F. Martin
840	International Capital Flows and U.S. Interest Rates	Francis E. Warnock Veronica C. Warnock
839	Effects of Financial Autarky and Integration: The Case of the South Africa Embargo	Brahima Coulibaly
838	General-to-specific Modeling: An Overview and Selected Bibliography	Julia Campos Neil R. Ericsson David F. Hendry
837	Currency Crashes and Bond Yields in Industrial Countries	Joseph E. Gagnon
836	Estimating Elasticities for U.S. Trade in Services	Jaime Marquez
835	SIGMA: A New Open Economy Model for Policy Analysis	Christopher Erceg Luca Guerrieri Christopher Gust
834	Optimal Fiscal and Monetary Policy with Sticky Wages and Sticky Prices	Sanjay K. Chugh
833	Exchange Rate Pass-through to U.S. Import Prices: Some New Evidence	Mario Marazzi Nathan Sheets Robert J. Vigufsson And Others
832	A Flexible Finite-Horizon Identification of Technology Shocks	Neville Francis Michael T. Owyang Jennifer E. Roush
831	Adjusting Chinese Bilateral Trade Data: How Big is China's Surplus	John W. Schindler Dustin H. Beckett
830	Order Flow and Exchange Rate Dynamics in Electronic Brokerage System Data	David W. Berger Alain P. Chaboud Sergey V. Chernenko Edward Howorka Raj S. Krishnasami Iyer David Liu Jonathan H. Wright